บทที่ 3 การวิเคราะห์และการนำเสนอข้อมูลเชิงปริมาณ

โดย

นายธเนศ วินัยรักษ์

การวิเคราะห์และนำเสนอข้อมูลเชิงปริมาณด้วยตารางความถี่

ขั้นตอนการเขียนตารางความถี่ของข้อมูลเชิงปริมาณที่มีข้อมูลทั้งหมดเป็นจำนวนเต็ม

1. กำหนดจำนวนอันตรภาคชั้นเป็น k ชั้น

2. กำหนดค่าเริ่มต้นและค่าสุดท้ายที่ครอบคลุมทุกค่าของข้อมูล โดยที่ค่าเริ่มต้นคือค่าต่ำสุดหรือค่าที่น้อยกว่าค่าต่ำสุดของข้อมูล และค่าสุดท้ายคือค่าสูงสุดหรือค่าที่มากกว่าค่าสูงสุดของข้อมูล

3. คำนวณความกว้างของอันตรภาคชั้น โดยหาได้จาก

$$\frac{\text{ค่าสุดท้าย} - \text{ค่าเริ่มต้น}}{\text{จำนวนอันตรภาคชั้น}}$$

ถ้าค่าที่คำนวณได้ไม่เป็นจำนวนเต็มให้ปัดเศษขึ้นเป็นจำนวนเต็มเสมอ

การวิเคราะห์และนำเสนอข้อมูลเชิงปริมาณด้วยตารางความถี่

4. กำหนดอันตรภาคชั้นโดยที่

- ชั้นแรกที่ค่าเริ่มต้นที่กำหนดในข้อ 2. ถึงจำนวนที่ได้จากการนำค่าเริ่มต้นที่กำหนดในข้อ 2. บวกกับความกว้างของอันตรภาคชั้นลบด้วย 1

- ชั้นที่สองมีค่าเริ่มต้นเป็นค่าสุดท้ายของชั้นแรกบวกด้วย 1 ถึงค่าเริ่มต้นของชั้นที่สองบวกกับความกว้างของอันตรภาคชั้นลบด้วย 1

- ทำเช่นนี้เรื่อย ๆ จนถึงชั้นที่ k (ในกรณีที่ $\frac{\text{ค่าสุดท้าย} - \text{ค่าเริ่มต้น}}{\text{จำนวนอันตรภาคชั้น}}$ เป็นจำนวนเต็ม ค่าสุดท้ายของชั้นที่ k จะไม่เท่ากับค่าสุดท้ายที่กำหนดในข้อ 2. แต่ต้องมากกว่าหรือเท่ากับข้อมูลทุกค่า

5. หาจำนวนข้อมูลทั้งหมดที่อยู่ในแต่ละอันตรภาคชั้น โดยทำรอยขีดแทนจำนวนไว้ในแต่ละอันตรภาคชั้น โดยปกติมักใช้รอยขีด 1 แทนหนึ่งค่า และเพื่อความสะดวกในการนับจำนวนข้อมูลที่อยู่ในแต่ละอันตรภาคชั้น เมื่อถึงทุก ๆ ข้อมูลที่ห้า มักนิยมทำรอยขีดแนวเฉียงหรือแนวนอนทับรอยขีดทั้งสี่ก่อนหน้านั้น ดังนี้ 𝍸 หรือ 𝍸

6. นับจำนวนข้อมูลจากรอยขีดที่ทำในข้อ 5. แล้วบันทึกจำนวนข้อมูลลงในช่องความถี่ของแต่ละอันตรภาคชั้น

ตัวอย่างที่ 1

คะแนนสอบวิชาคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ห้องหนึ่ง จำนวน 30 คน โดยคะแนนเป็นจำนวนเต็มเท่านั้น แสดงได้ดังนี้

85	112	112	123	109	85	87	123	111	112
87	126	94	92	93	91	99	121	122	128
102	118	88	106	111	98	128	94	95	129

จงเขียนตารางความถี่ ความถี่สัมพันธ์ และความถี่สะสมสัมพัทธ์ โดยสร้างตารางให้มีจำนวนอันตรภาคชั้น 5 ชั้น

ตัวอย่างที่ 1

เขียนตารางความถี่ดังนี้

1. กำหนดจำนวนอันตรภาคชั้นทั้งหมด 5

2. กำหนดค่าเริ่มต้นเท่ากับคะแนนต่ำสุดซึ่งคือ 85 คะแนน และค่าสุดท้ายเท่ากับคะแนนสูงสุดซึ่งคือ 120 คะแนน

3. คำนวณความกว้างของอันตรภาคชั้น ได้ดังนี้

$$\frac{\text{ค่าสุดท้าย} - \text{ค่าเริ่มต้น}}{\text{จำนวนอันตรภาคชั้น}} = \frac{129 - 85}{5} = 8.8$$

ดังนั้นความกว้างของอันตรภาคชั้นคือ 9 คะแนน

ตัวอย่างที่ 1

4. กำหนดอันตรภาคชั้น ได้ดังนี้

อันตรภาคชั้น	รอยชีด	ความถี่	ความถี่สะสม	ความถี่สัมพัทธ์		ความถี่สะสมสัมพัทธ์	
				สัดส่วน	ร้อยละ	สัดส่วน	ร้อยละ
85 – 93							
94 – 102							
103 – 111							
112 – 120							
121 - 129							

ตัวอย่างที่ 1

5. ให้นำข้อมูลเติมในตารางให้ครบถ้วน

อันตรภาคชั้น	ความถี่	ความถี่สะสม	ความถี่สัมพัทธ์		ความถี่สะสมสัมพัทธ์	
			สัดส่วน	ร้อยละ	สัดส่วน	ร้อยละ
85 – 93	8	8	$\frac{8}{30} \approx 0.27$	27	0.27	27
94 – 102	6	14	$\frac{6}{30} = 0.20$	20	0.47	47
103 – 111	4	18	$\frac{4}{30} \approx 0.13$	13	0.50	50
112 – 120	4	22	$\frac{4}{30} \approx 0.13$	13	0.63	63
121 - 129	8	30	$\frac{8}{30} \approx 0.27$	27	1.00	100

การวิเคราะห์และนำเสนอข้อมูลเชิงปริมาณด้วยตารางความถี่

- 1. ถ้ามีข้อมูลค่าแตกต่างในจากค่าอื่น ๆ ในข้อมูลนั้นมาก ควรกำหนดอันตรภาคชั้นเปิด เช่น

อันตรภาคชั้น (คะแนน)	ความถี่
น้อยกว่า 40	1
40 – 49	2
50 – 59	6
60 – 69	20
70 – 79	21
80 – 89	8
90 - 99	2

การวิเคราะห์และนำเสนอข้อมูลเชิงปริมาณด้วยตารางความถี่

- 2. ควรกว้างของแต่ละอันตรภาคชั้นไม่จำเป็นต้องเท่ากัน ทั้งนี้ขึ้นอยู่กับวัตถุประสงค์ของการใช้ข้อมูล เช่น

อันตรภาคชั้น (คะแนน)	ความถี่
80 ขึ้นไป	4
70 – 79	2
60 – 69	7
50 – 59	9
0 - 49	1

ในกรณีที่ข้อมูลไม่เป็นจำนวนเต็ม การกำนดอันตรภาคชั้นอาจกำหนดให้อยู่ในรูปช่วง ดังแสดงในตัวอย่างต่อไปนี้

ตัวอย่างที่ 2

จากการสำรวจชั่วโมงเฉลี่ยของการทำงานใน 1 สัปดาห์ ของคนงงานในโรงงาน 8 แห่ง ในแต่ละเขตอุตสาหกรรม ได้ข้อมูลดังนี้

จำนวนชั่วโมงเฉลี่ยของการทำงานต่อคนต่อสัปดาห์ในแต่ละโรงงาน					
เขต ก	**ชั่วโมง**	**เขต ข**	**ชั่วโมง**	**เขต ค**	**ชั่วโมง**
คนงาน ก1	35.0	คนงาน ข1	40.0	คนงาน ค1	43.0
คนงาน ก2	48.0	คนงาน ข2	50.0	คนงาน ค2	48.8
คนงาน ก3	45.0	คนงาน ข3	35.4	คนงาน ค3	43.3
คนงาน ก4	43.0	คนงาน ข4	38.8	คนงาน ค4	53.1
คนงาน ก5	38.2	คนงาน ข5	40.2	คนงาน ค5	35.6
คนงาน ก6	50.0	คนงาน ข6	45.0	คนงาน ค6	41.1
คนงาน ก7	39.8	คนงาน ข7	45.0	คนงาน ค7	34.8
คนงาน ก8	40.7	คนงาน ข8	40.0	คนงาน ค8	51.6

จงหาสร้างตารางความถี่ให้มีจำนวนอันตรภาคชั้นเท่ากับ 5

ตัวอย่างที่ 2

เขียนตารางความถี่ดังนี้

1. กำหนดจำนวนอันตรภาคชั้นทั้งหมด 5

2. เนื่องจากข้อมูลมีค่าต่ำสุดคือ 34.8 ชั่วโมงต่อสัปดาห์ และค่าสูงสุดคือ 53.1 ชั่วโมงต่อสัปดาห์ จึงกำหนดค่าเริ่มต้นคือ 30 ชั่วโมงต่อสัปดาห์ และค่าสุดท้ายคือ 55 ชั่วโมงต่อสัปดาห์

3. คำนวณความกว้างของอันตรภาคชั้น ได้ดังนี้

$$\frac{\text{ค่าสุดท้าย} - \text{ค่าเริ่มต้น}}{\text{จำนวนอันตรภาคชั้น}} = \frac{55 - 30}{5} = 5$$

ดังนั้นความกว้างของอันตรภาคชั้นคือ 5 คะแนน

ตัวอย่างที่ 2

4. กำหนดอันตรภาคชั้นในรูปช่วงโดยแบ่งเป็น 5 ชั้น ได้ดังนี้

จำนวนชั่วโมง (x)	รอยขีด	ความถี่
$30 \leq x < 35$		
$35 \leq x < 40$		
$40 \leq x < 45$		
$45 \leq x < 50$		
$50 \leq x < 55$		

ตัวอย่างที่ 2

5. ให้นำข้อมูลเติมในตารางให้ครบถ้วน

จำนวนชั่วโมง (x)	ความถี่
$30 \leq x < 35$	1
$35 \leq x < 40$	6
$40 \leq x < 45$	8
$45 \leq x < 50$	5
$50 \leq x < 55$	4

การวิเคราะห์และนำเสนอข้อมูลเชิงปริมาณด้วยตารางความถี่

- การกำหนดจำนวนอันตรภาคชั้นในการเขียนตารางความถี่ไม่มีกฎเกณฑ์แน่นอนตายตัวขึ้นอยู่กับลักษณะการกระจายของข้อมูลหรือความแตกต่างระหว่างข้อมูลว่ามีมากน้อยเพียงใด และอาจขึ้นอยู่กับรายละเอียดของข้อมูลที่ต้องการทราบด้วย เช่น ถ้าข้อมูลมีมีความแตกต่างกันมาก มักจะกำหนดจำนวนอันตรภาคชั้นให้น้อย เพื่อไม่ให้อันตรภาคชั้นส่วนใหญ่มีความถี่เป็นศูนย์ เนื่องจากไม่มีข้อมูลใดอยู่ในอันตรภาคชั้นนั้นเลย หรือถ้าต้องการทราบรายละเอียดเกี่ยวกับข้อมูลนั้นมาก ก็ควรกำหนดจำนวนอันตรภาคชั้นให้มากขึ้น โดยทั่วไปจำนวนอันตรภาคชั้นที่นิยมใช้มักจะอยู่ระหว่าง 7 ถึง 15 ชั้น หรือส่วนใหญ่จะไม่ต่ำกว่า 5 ชั้น

- **หมายเหตุ** สำหรับข้อมูลที่ไม่เป็นจำนวนเต็มอาจกำหนดช่วงของอันตรภาคชั้นในรูป $a < x \leq b$ แต่ในที่นี้จะเขียนในรูป $a \leq x < b$ เท่านั้น

แบบฝึกหัด 3.1

1. จำนวนเซลล์เม็ดเลือดขาว (เซลล์ต่อเลือด 1 ลูกบาศก์มิลลิเมตร) ของครูระดับมัธยมศึกษาตอนปลายฝนนภาคตะวันออกเฉียงเหนือจำนวน 40 คน แสดงได้ดังนี้

9328	8100	9517	8909	9161	9015	8848	9612	8182	9107
9006	9207	9418	8708	9247	8594	8820	8401	8409	8701
8767	8969	9297	8985	8735	9313	9419	9123	8716	8501
8974	8412	9345	8112	9432	9211	9078	8988	8811	8841

จงเขียนตารางความถี่ของข้อมูลชุดนี้ โดยกำหนดให้จำนวนอันตรภาคชั้นเท่ากับ 9 ชั้น ค่าเริ่มต้นและค่าสุดท้าย

แบบฝึกหัด 3.1

2. จำนวนประชากรที่อาศัยอยู่ในหมู่บ้านต่าง ๆ ของอำเภอหนึ่ง จำนวน 66 หมู่บ้าน แสดงได้ดังนี้

345	494	604	730	468	607	395	494	530	563
515	505	468	461	560	490	556	444	523	549
624	439	384	402	354	629	545	472	582	596
585	523	593	439	505	475	574	417	629	420
574	486	527	611	578	490	490	585	516	505
431	541	523	523	468	549	479	604	464	691
575	585	603	417	446	406				

1) จงเขียนตารางความถี่ของข้อมูลชุดนี้ โดยกำหนดให้ 341 - 380 เป็นอันตรภาคชั้นแรก และแต่ละอันตรภาคชั้นมีความกว้างเท่ากัน

2) มีหมู่บ้านที่มีประชากรอาศัยอยู่ต่ำกว่า 501 คน มีกี่หมู่บ้าน

3) มีหมู่บ้านที่มีประชากรอาศัยอยู่ตั้งแต่ 381 ถึง 580 คน กี่หมู่บ้าน

4) จำนวนหมู่บ้านที่มีประชากรอาศัยมากกว่า 660 คน คิดเป็นร้อยละเท่าใดของจำนวนหมู่บ้านทั้งหมด

การวิเคราะห์และนำเสนอข้อมูลเชิงปริมาณด้วยแผนภาพ

- ประกอบด้วย
  - 1. ฮิสโทแกรม
  - 2. แผนภาพจุด
  - 3. แผนภาพลำต้นและใบ
  - 4. แผนภาพกล่อง
  - 5. แผนภาพการกระจาย

ฮิสโทแกรม

- ฮิสโตแกรม (histogram) เป็นการนำเสนอข้อมูลเชิงปริมาณที่สร้างชจากตารางความถี่ โดยใช้แท่งสี่เหลี่ยมมุฉากที่รเรียงตดิกันบนแกนนอน เมื่อแกนนอนแทนค่าของข้อมูล ความสูงของแท่งสี่เหลี่ยมมุมฉากจะปสดงความถี่ของข้อมูล ซึ่งการแสดงความถี่ของข้อมูลอาจนำเสนอความถี่ของข้อมูลเพียงค่าเดียวหรือข้อมูลในแต่ละอันตรภาคชั้น โดยความกว้างของแท่งสี่เหลี่ยมแต่ละแท่งจะสอดคล้องกับความกว้างของแต่ละอันตรภาคชั้นของตารางความถี่

ฮิสโทแกรม

พิจารณาจากสถานการณ์ต่อไปนี้

คณะกรรมการหมู่บ้านจัดสรรแห่งหนึ่งสำรวจข้อมูลจำนวนสมาชิกในแต่ละครอบครัว จำนวน 20 ครอบครัว โดยเรียงข้อมูลจากน้อยไปมากได้ดังนี้

2	2	2	3	3	3	3	3	4	4
4	4	4	4	5	5	6	6	6	6

จากข้อมูลข้างต้นสามารถเขียนตารางความถี่ได้ดังนี้

ฮิสโทแกรม

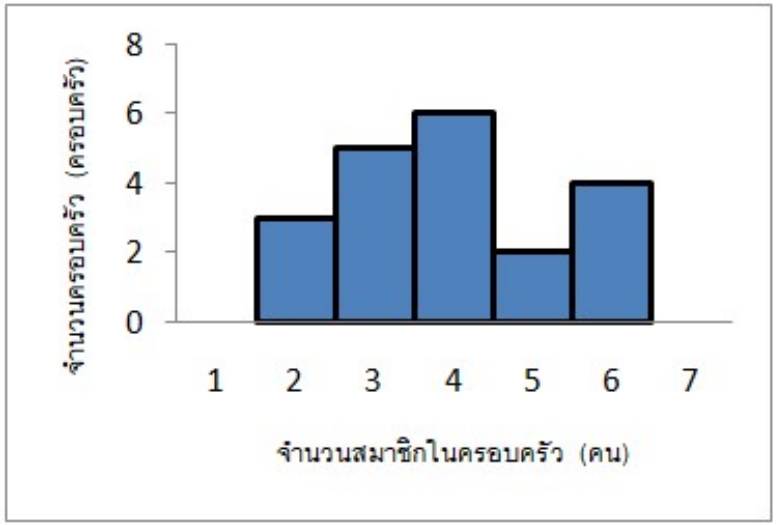
จำนวนสมาชิกในครอบครัว (คน)	จำนวนครอบครัว (ครอบครัว)
2	
3	
4	
5	
6	
รวม	

ฮิสโทแกรม

จำนวนสมาชิกในครอบครัว (คน)	จำนวนครอบครัว (ครอบครัว)
2	3
3	5
4	6
5	2
6	4
รวม	20

ฮิสโทแกรม

สามารถนำเสนอข้อมูลข้างต้นโดยใช้ฮิสโทแกรมได้ดังนี้



จากฮิสโทแกรม จะเห็นว่าครอบครัวที่มีสมาชิก 4 คน มีจำนวนมากที่สุด ไม่มีครอบครัวใดและเลยที่มีสมาชิกมากกว่า 6 คน และไม่มีลครอบครัวใดเลลที่มีสมาชิกเพียง 1 คน

ฮิสโทแกรม

- **ข้อสังเกต**
  - 1. จุดกึ่งกลางของฐานของแต่ละแท่งสี่เหลี่ยมมุมฉากแทนข้อมูลแต่ละค่า โดยในที่นี้คือจำนวนสมาชิกในครอบครัว ซึ่งได้แก่ 2, 3, 4, 5 และ 6 คน
  - 2. ความสูงของแต่ละแท่งสี่เหลี่ยมมุมฉากแทนความถี่ของข้อมูลแต่ละค่า โดยในที่นี้คือจำนวนครอบครัว ซึ่งได้แก่ 3, 5, 6, 2 และ 4 ครอบครัว
  - 3. ฮิสโทแกรมมีลักษณะคล้ายแผนภูมิแท่ง แต่แผนภูมิแท่งใช้นำเสนอข้อมูลเชิงคุณภาพ

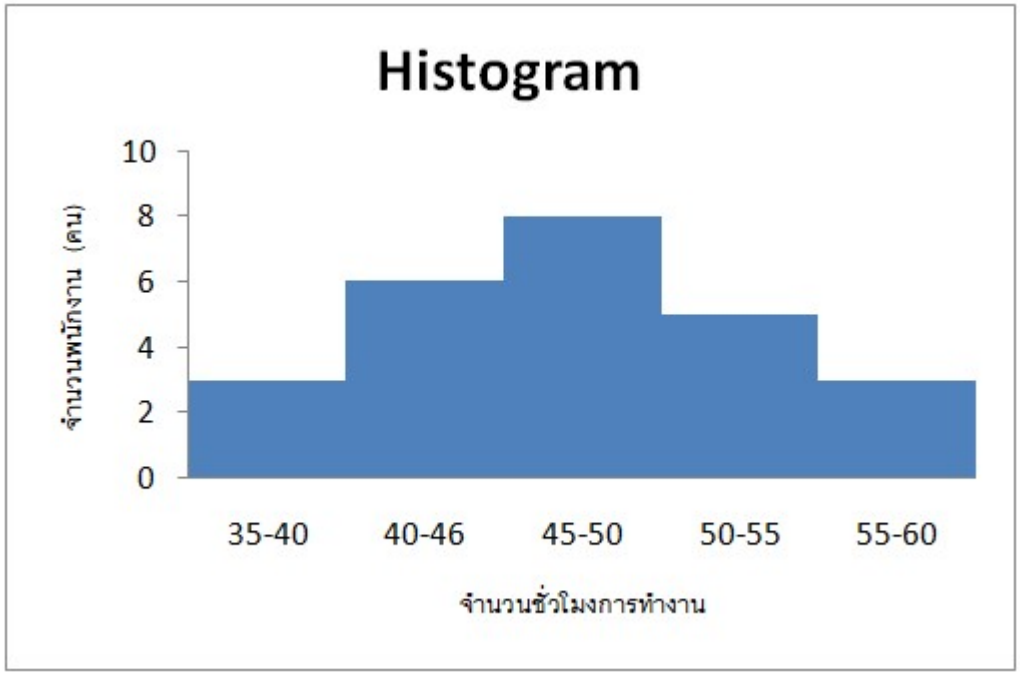
ฮิสโทแกรม

พิจารณาสถานการณ์ต่อไปนี้

ฝ่ายทรัพยากรบุคลลของหน่วยงานแห่งหนึ่งได้เก็บข้อมูลจำนวนชั่วโมงการทำงานในหนึ่งสัปดาห์ของพนักงาน จำนวน 25 คน ได้ข้อมูลดังตารางความถี่ต่อไปนี้

จำนวนชั่วโมงการทำงาน (x)	จำนวนพนักงาน (คน)
$35 \le x < 40$	3
$40 \le x < 45$	6
$45 \le x < 50$	8
$50 \le x < 55$	5
$55 \le x < 60$	3
รวม	25

ฮิสโทแกรม



- จากฮิสโทแกรม จะเห็นว่าในหนึ่งสัปดาห์ พนักงานที่มีจำนวนชั่วโมงการทำงานตั้งแต่ 45 ชั่วโมง แต่น้อยกว่า 50 ชั่วโมง มีจำนวนมากที่สุด และจำนวนพนักงานที่มีชั่วโมงในการทำงานตั้งแต่ 35 ชั่วโมง แต่น้อยกว่า 40 ชั่วโมง เท่ากับจำนวนพนักงานที่มีจำนวนชั่วโมงการทำงานตั้งแต่ 55 ชั่วโมง แต่น้อยกว่า 60 ชั่วโมง

ฮิสโทแกรม

- **ขอบล่างของชั้น (lower class boundary) คือ ค่ากึ่งกลางระหว่างค่าของข้อมูลที่มากที่สุดในชั้นก่อนหน้ากับค่าของข้อมูลที่น้อยที่สุดในชั้นนั้น**

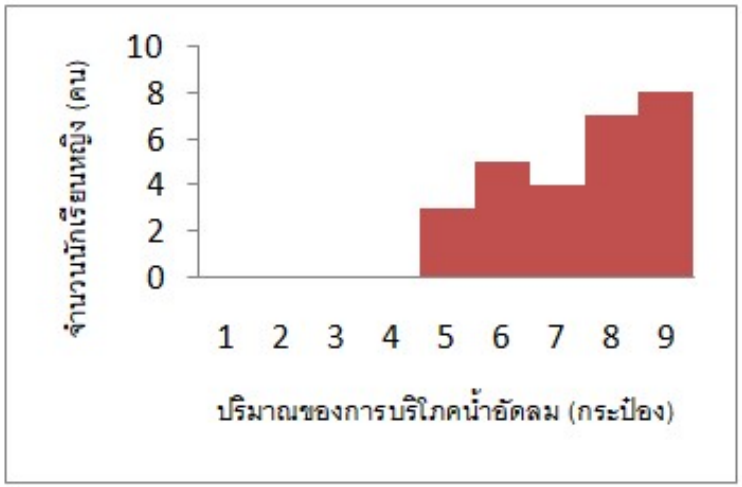
- **ขอบบนของชั้น (upper class boundary) คือ ค่ากึ่งกลางระหว่างค่าของข้อมูลที่มากที่สุดในชั้นนั้นกับค่าของข้อมูลที่น้อยที่สุดในชั้นถัดไป**

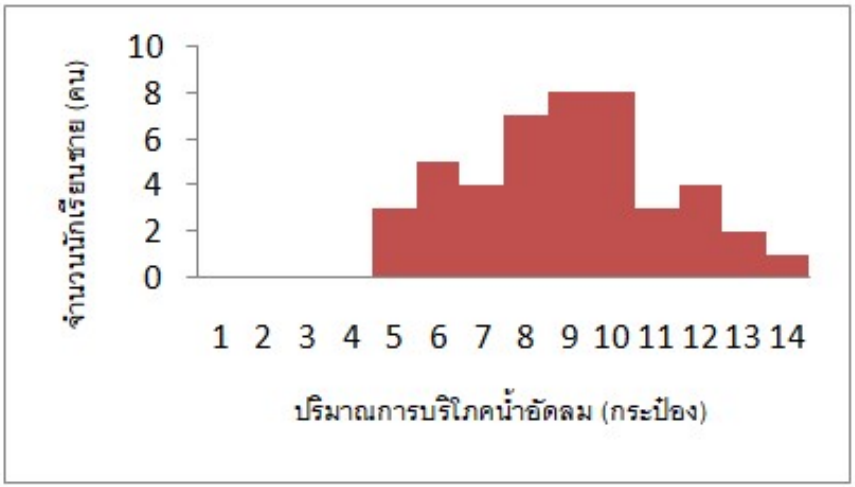
หมายเหตุ

การหาขอบล่างของชั้นของอันตรภาคชั้นแรกให้พิจารณาเสมือนว่ามีอันตรภาคชั้นที่ต่ำกว่าอันตรภาคชั้นนี้อีกหนึ่งชั้น ในทำนองเดียวกัน การหาค่าขอบบนของชั้นในอันตรภาคชั้นสุดท้ายให้พิจารณาเสมือนว่ามีอันตรภาคชั้นที่สูงกว่าอันตรภาคชั้นนี้อีกหนึ่งชั้น

ตัวอย่างที่ 3

ปริมาณน้ำอัดลม (กระป๋อง) ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ในโรงเรียนแห่งหนึ่งในหนึ่งสัปดาห์ แสดงด้วยฮิสโทแกรมได้ดังนี้





1) นักเรียนหญิงจำนวนมากที่สุดบริโภคน้ำอัดลมกี่กระป๋องในหนึ่งสัปดาห์

2) นักเรียนชายบริโภคน้ำอัดลมมากที่สุดกี่กระป๋องในหนึ่งสัปดาห์

3) จงสรุปลักษณะที่น่าสนใจเกี่ยวกับพฤติกรรมการบริโภคน้ำอัดลมของนักเรียนหญิงและนักเรียนชายในโรงเรียนแห่งนี้

ตัวอย่างที่ 3

- 1) จากฮิสโทแกรมแสดงปริมาณการบริโภคน้ำอัดลมของนักเรียนหญิงในโรงเรียนแห่งนี้ จะได้ว่านักเรียนหญิงบริโภคน้ำอัดลม 2 กระป๋องในหนึ่งสัปดาห์มีจำนวนมากที่สุด

- 2) จากฮิสโทแกรมแสดงปริมาณการบริโภคน้ำอัดลมของนักเรียนชายในโรงเรียนแห่งนี้ จะได้ว่านักเรียนชายบริโภคน้ำอัดลม 14 กระป๋องในหนึ่งสัปดาห์มีจำนวนมากที่สุด

- 3) จากฮิสโทแกรม สรุปได้ว่าในหนึ่งสัปดาห์ นักเรียนหญิงส่วนใหญ่บริโภคน้ำอัดลมน้อยกว่า 6 กระป๋อง ในขณะที่นักเรียนชายส่วนใหญ่บริโภคน้ำอัดลมมากกว่า 6 กระป๋อง

แผนภาพจุด

- แผพภาพจุด (dot plot) เป็นการนำเสนอข้อมูลเชิงปริมาณโดยใช้จุดหรือวงกลมเล็ก ๆ แทนข้อมูลแต่ละตัว เขียนเรียงไว้เหนือเส้นในแนวนอนที่มีสเกล จุดหรือวงกลมเล็ก ๆ ดังกล่วจะเรียงกันในแนวตั้งตรงกับตำแหน่งซึ่งแสดงค่าของข้อมูลแต่ละตัว

พิจารณาสถานการณ์ต่อไปนี้

จากการแข่งขันกีฬาโอลิมปิกฤดูหนาว 2018 ซึ่งจัดขึ้นที่สาธารณรัฐเกาหลี ใน พ.ศ. 2561 สามารถแสดงข้อมูลจำนวนเหรียญทองของประเทศที่ได้เหรียญทองจำนวน 22 ประเทศ โดยเรียงข้อมูลจากน้ยไปมากได้ดังนี้

14	14	11	9	8	7	5	5	5	5	4
3	2	2	2	1	1	1	1	1	1	1

แผนภาพจุด

จากข้อมูลข้างต้น เขียนตารางความถี่ได้ดังนี้

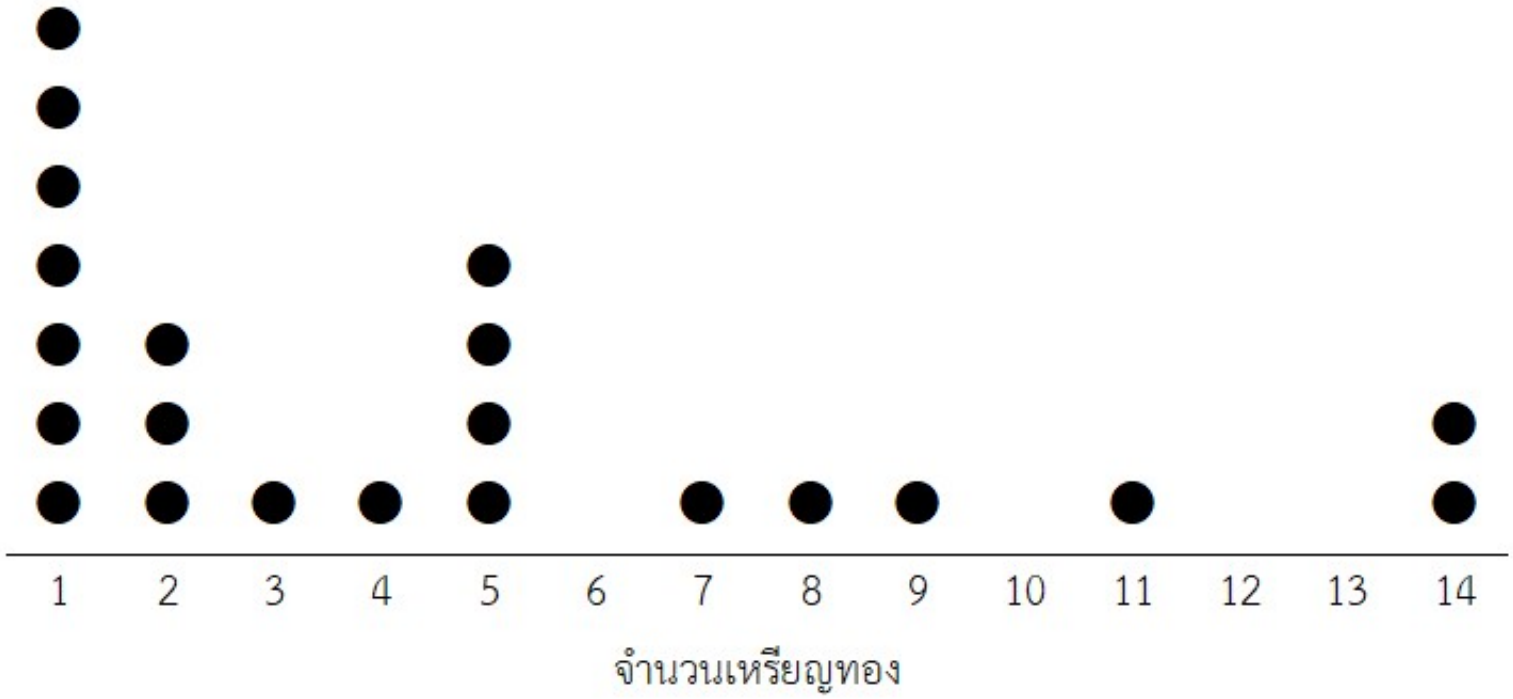
จำนวนเหรียญทอง (เหรียญ)	จำนวนประเทศ
1	
2	
3	
4	
5	
7	
8	
9	
11	
14	

แผนภาพจุด

จำนวนเหรียญทอง (เหรียญ)	จำนวนประเทศ
1	7
2	3
3	1
4	1
5	4
7	1
8	1
9	1
11	1
14	2

แผนภาพจุด

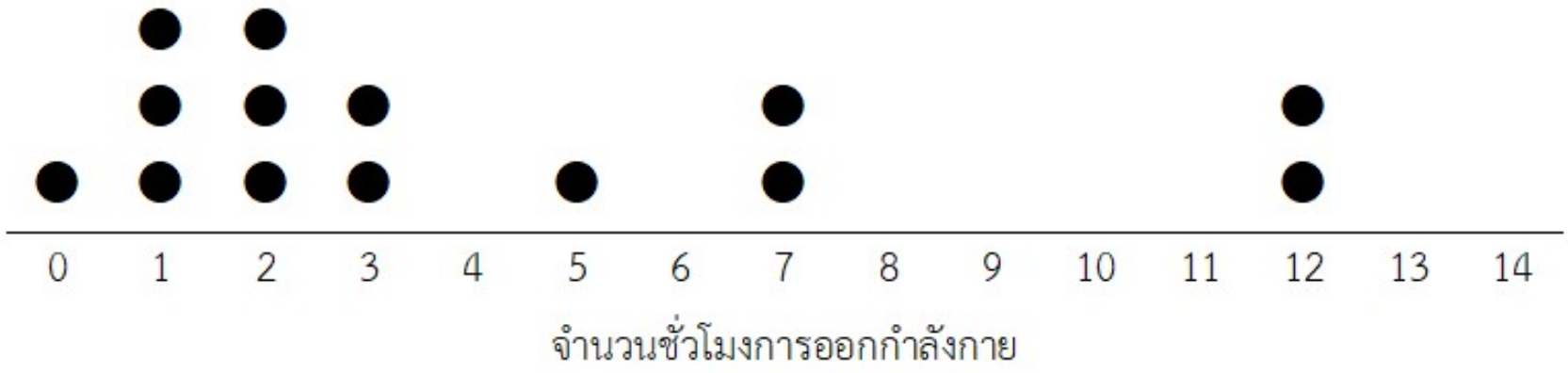
จากตารางความถี่ข้างต้น สามารถเขียนแผนภาพจุดได้ดังนี้



จากแผนภาพจุด จะเห็นว่าแระเทศที่ได้เหรียญทอง 1 เหรียญ มีจำนวนมากที่สุด และมีเพียง 3 ประเทศที่ได้เหรียญทองมากกว่า 10 เหรียญ

ตัวอย่างที่ 4

จำนวนชั่วโมงการออกกำลังกายของครูในโรงเรียนแห่งหนึ่งในหนึ่งสัปดาห์ แสดงด้วยแผนภาพจุดได้ดังนี้



จงสรุปลักษณะที่น่าสนใจเกี่ยวกับจำนวนชั่วโมงการออกกำลังกายของครูในโรงเรียนแห่งนี้ในหนึ่งสัปดาห์

- จากแผนภาพจุด สรุปได้ว่า ในหนึ่งสัปดาห์ครูส่วนใหญ่ออกกำลังกายน้อยกว่า 4 ชั่วโมง และมีครูจำนวน 2 คน ที่ออกกำลังกายมากถึง 12 ชั่วโมงต่อสัปดาห์

แผนภาพลำต้นและใบ

- แผนภาพลำต้นและใบ (stem and leaf plot) เป็นการนำเสนอข้อมูลเชิงปริมาณโดยใช้แผนภาพที่มีการแสดงข้อมูลโดยเรียงลำดับจากน้อยไปมากและแบ่งการแสดงข้อมูลออกเป็นสองส่วนที่เรียกว่า ส่วนลำต้น และส่วนใบ ใบที่นี้กำหนดส่วนใบเป็นเลขโดดในหลักหน่วย และตัวเลขที่เหลือเป็นส่วนลำต้น เช่น 298 จะมี 29 เป็นส่วนลำต้น และ 8 เป็นส่วนใบ

แผนภาพลำต้นและใบ

พิจารณาสถานการณ์ต่อไปนี้

ร้านอาหารแห่งหนึ่งสำรวจอายุ (ปี) ของผู้ที่มาใช้บริการที่ร้านในหนึ่งวัน ได้ข้อมูลดังนี้

1	27	2	20	27	23	31	30	9	29
31	8	28	25	26	40	37	23	34	49
52	31	1	4	5	58	28	57	31	32
3	4	25	31	29	57	44	2	35	24
4	30	56	63	48					

เรียงข้อมูลจากน้อยไปมากได้ดังนี้

แผนภาพลำต้นและใบ

1	1	2	2	3	4	4	4	5	8
9	20	23	23	24	25	25	26	27	27
28	28	29	29	30	30	31	31	31	31
31	32	34	35	37	40	44	48	49	52
56	57	57	58	63					

แผนภาพลำต้นและใบ

จากข้อมูลข้างต้น สามารถเขียนแผนภาพลำต้นและใบได้ดังนี้

0	
1	
2	
3	
4	
5	
6	

แผนภาพลำต้นและใบ

0	1	1	2	2	3	4	4	4	5	8	9		
1													
2	0	3	3	4	5	5	6	7	7	8	8	9	9
3	0	0	1	1	1	1	1	2	4	5	7		
4	0	4	8	9									
5	2	6	7	7	8								
6	3												

- จากแผนภาพลำต้นและใบ จะเห็นว่าในวันที่สำรวจข้อมูล ผู้ที่มาใช้บริการที่ร้านอาหารแห่งนี้มีอายุมากที่สุด 63 ปี และมีอายุน้อยที่สุด 1 ปี โดยที่ผู้ที่มาใช้บริการที่ร้านอาหารแห่งนี้อยู่ในช่วงอายุ 20 – 29 ปีมากที่สุดเช่น

แผนภาพลำต้นและใบ

- **ข้อสังเกต**
  - ตัวเลขในส่วนลำต้นจะต้องเขียนเรียงลำดับให้ครบทุกตัว แต่ไม่ต้องเขียนตัวเลขในส่วนใบหากไม่มีข้อมูลนั้น เช่น จากสถานการณ์ขางต้น จะเห็นว่าไม่มีข้อมูลในช่วงอายุ 10 – 19 ปี แต่ยังคงปรากฏ 1 ใบส่วนลำต้น

- นอกจากนี้สามารถเขียนแผนภาพลำต้นและใบในการนำเสนอข้อมูลสองชุดในแผนภาพเดียวกันได้โดยใช้ส่วนลำต้นร่วมกัน เพื่อใช้เปรียบเทียบข้อมูลทั้งสองชุด

แผนภาพลำต้นและใบ

พิจารณาสถานการณ์ต่อไปนี้

คะแนนสอบวิชาภาษาอังกฤษของนักเรียนชั้นมัธยมศึกษาปีที่ 6/1 และ 6/2 ในโรงเรียนขนาดเล็กแห่งหนึ่ง ซึ่งมีคะแนนเต็ม 100 คะแนน โดยเรียงข้อมูลจากน้อยไปมาก แสดงได้ดังนี้

นักเรียนชั้นมัธยมศึกษาปีที่ 6/1

56	60	65	69	73	73	75	78	83	85
87	87	89	93	94	97				

นักเรียนชั้นมัธยมศึกษาปีที่ 6/2

57	63	64	65	68	68	70	72	73	73
75	78	81	83	87	98				

จากข้อมูลข้างต้น สามารถเขียนแผนภาพลำต้นและใบได้ดังนี้

แผนภาพลำต้นและใบ

จากข้อมูลข้างต้น สามารถเขียนแผนภาพลำต้นและใบได้ดังนี้

นักเรียนชั้นมัธยมศึกษาปีที่ 6/1		นักเรียนชั้นมัธยมศึกษาปีที่ 6/2
	5	
	6	
	7	
	8	
	9	

แผนภาพลำต้นและใบ

นักเรียนชั้นมัธยมศึกษาปีที่ 6/1						นักเรียนชั้นมัธยมศึกษาปีที่ 6/2						
				6	5	7						
		9	5	0	6	3	4	5	8	8		
	8	5	3	3	7	0	2	2	3	3	5	8
9	7	7	5	3	8	1	3	7				
					9	8						

- จากแผนภาพลำต้นและใบ
  - คะแนนสอบสูงสุดวิชาภาษาอังกฤษของนักเรียนชั้นมัธยมศึกษาปีที่ 6/1 และ 6/2 คือ 97 และ 98 คะแนน ตามลำดับ

แผนภาพลำต้นและใบ

นักเรียนชั้นมัธยมศึกษาปีที่ 6/1						นักเรียนชั้นมัธยมศึกษาปีที่ 6/2						
				6	5	7						
		9	5	0	6	3	4	5	8	8		
	8	5	3	3	7	0	2	2	3	3	5	8
9	7	7	5	3	8	1	3	7				
					9	8						

- คะแนนสอบต่ำสุดวิชาภาษาอังกฤษของนักเรียนชั้นมัธยมศึกษาปีที่ 6/1 และ 6/2 คือ 56 และ 57 คะแนน ตามลำดับ

แผนภาพลำต้นและใบ

นักเรียนชั้นมัธยมศึกษาปีที่ 6/1						นักเรียนชั้นมัธยมศึกษาปีที่ 6/2						
				6	5	7						
		9	5	0	6	3	4	5	8	8		
	8	5	3	3	7	0	2	2	3	3	5	8
9	7	7	5	3	8	1	3	7				
					9	8						

- จำนวนนักเรียนที่ได้คะแนนสอบวิชาภาษาอังกฤษตั้งแต่ 80 คะแนนขึ้นไป ของนักเรียนชั้นมัธยมศึกษาปีที่ 6/1 มากกว่าชั้นมัธยมศึกษาปีที่ 6/2

แผนภาพกล่อง

- แผนภาพกล่อง (box plot) เป็นการนำเสนอข้อมูลเชิงปริมาณที่แสดงตำแหน่งสำคัญของข้อมูลซึ่งประกอบด้วยค่าต่ำสุด ค่าสูงสุด และควอร์ไทล์ (quartile) นอกจากนี้ แผนภาพกล่องสามารถใช้ในการตรวจสอบว่ามีข้อมูลส่วนใหญ่หรือไม่ โดยจะเรียกข้อมูลดังกล่าวว่า **ค่านอกเกณฑ์ (outlier)**

แผนภาพกล่อง

1. เรียงข้อมูลจากน้อยไปมาก จากนั้นหาค่าต่ำสุดและค่าสูงสุดของข้อมูล
2. หาควอร์ไทล์ที่ 1 (Q_1) ควอร์ไทล์ที่ 2 (Q_2) และควอร์ไทล์ที่ 3 (Q_3) โดยที่

- ควอร์ไทล์ที่ 1 (Q_1) คือค่าที่มีจำนวนข้อมูลที่มีค่าน้อยกว่าค่านี้อยู่ประมาณหนึ่งในสี่ของจำนวนข้อมูลทั้งหมด
- ควอร์ไทล์ที่ 2 (Q_2) คือค่าที่มีจำนวนข้อมูลที่มีค่าน้อยกว่าค่านี้อยู่ประมาณครึ่งหนึ่งของจำนวนข้อมูลทั้งหมด
- ควอร์ไทล์ที่ 3 (Q_3) คือค่าที่มีจำนวนข้อมูลที่มีค่าน้อยกว่าค่านี้อยู่ประมาณสามในสี่ของจำนวนข้อมูลทั้งหมด

แผนภาพกล่อง

ถ้า n เป็นจำนวนข้อมูลทั้งหมด สามารถหาตำแหน่งของควอร์ไทล์ได้ดังนี้

Q_1 **อยู่ในตำแหน่งที่** $\frac{n+1}{4}$

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

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

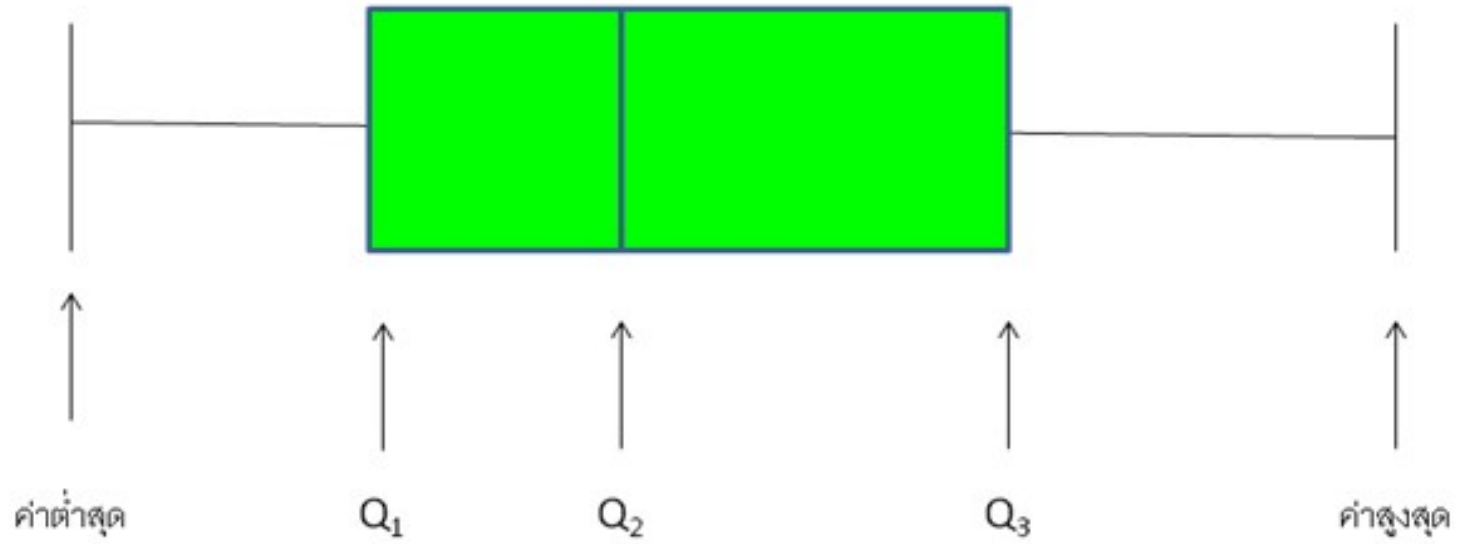
3. หาค่า $Q_1 - 1.5(Q_3 - Q_1)$ และ $Q_3 + 1.5(Q_3 - Q_1)$

4. พิจารณาว่าชุดข้อมูลมีค่านอกเกณฑ์หรือไม่ โดยในที่นี้ค่านอกเกณฑ์คือข้อมูลที่มีค่าน้อยกว่า $Q_1 - 1.5(Q_3 - Q_1)$ หรือข้อมูลที่มีค่ามากกว่า $Q_3 + 1.5(Q_3 - Q_1)$

แผนภาพกล่อง

5. ตัวอย่างการเขียนแผนภาพกล่อง

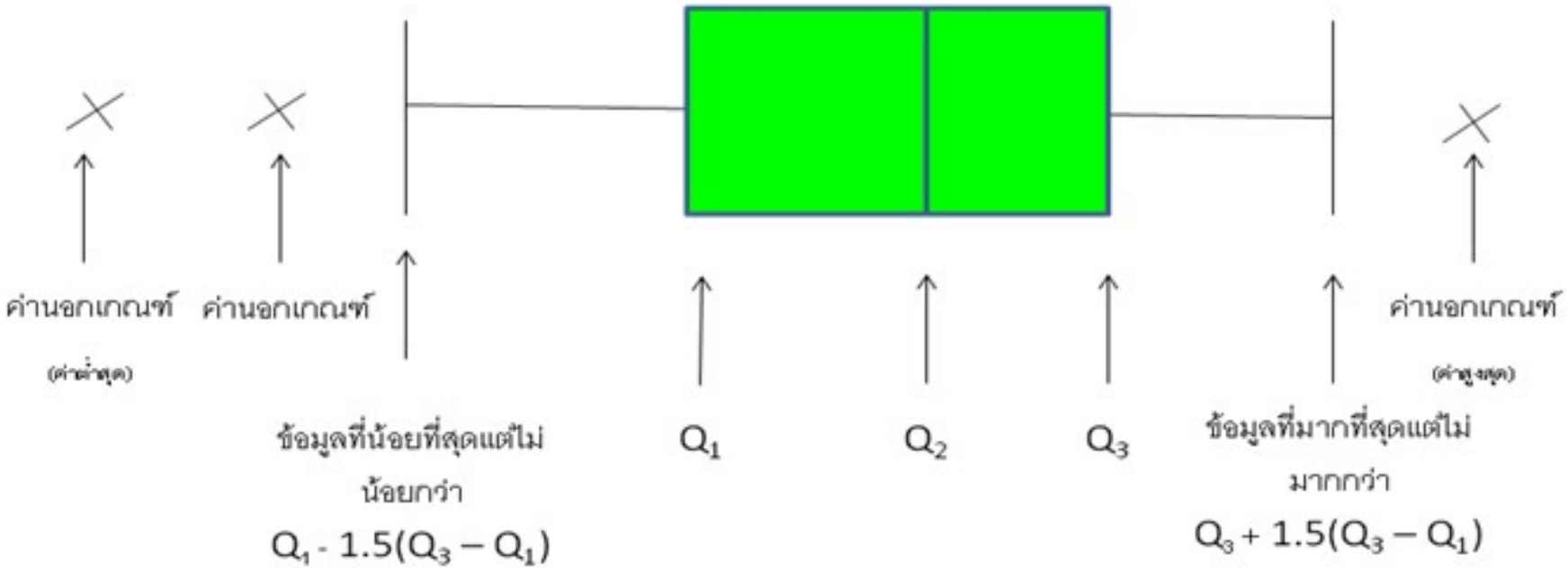
- กรณีที่ไม่มีค่านอกเกณฑ์



แผนภาพกล่อง

5. ตัวอย่างการเขียนแผนภาพกล่อง

- กรณีที่มีค่านอกเกณฑ์



ตัวอย่างที่ 5

คะแนนสอบวิชาภาษาไทยของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ห้องหนึ่ง จำนวน 27 คน ซึ่งมีคะแนนเต็ม 100 คะแนน โดยเรียงข้อมูลจากน้อยไปมาก แสดงได้ดังนี้

59	60	61	63	65	66	66	66	66	69
69	70	71	72	72	75	75	75	76	79
81	88	88	89	90	82	97			

จงเขียนผนภาพกล่องเพื่อนำเสนอข้อมูลชุดนี้

ตัวอย่างที่ 5

จากโจทย์ มีข้อมูลทั้งหมด 27 ตัว เขียนแผนภาพกล่องได้ตามขั้นตอนต่อไปนี้

1. หาค่าต่ำสุดของข้อมูล ซึ่งคือ 59 และค่าสูงสุดของข้อมูล ซึ่งคือ 97

2. หา Q_1, Q_2 และ Q_3 ได้ดังนี้

$$Q_1 \text{ อยู่ในตำแหน่งที่ } \frac{27+1}{4} = 7 \text{ ดังนั้น } Q_1 = 66$$

$$Q_2 \text{ อยู่ในตำแหน่งที่ } \frac{2(27+1)}{4} = 14 \text{ ดังนั้น } Q_2 = 72$$

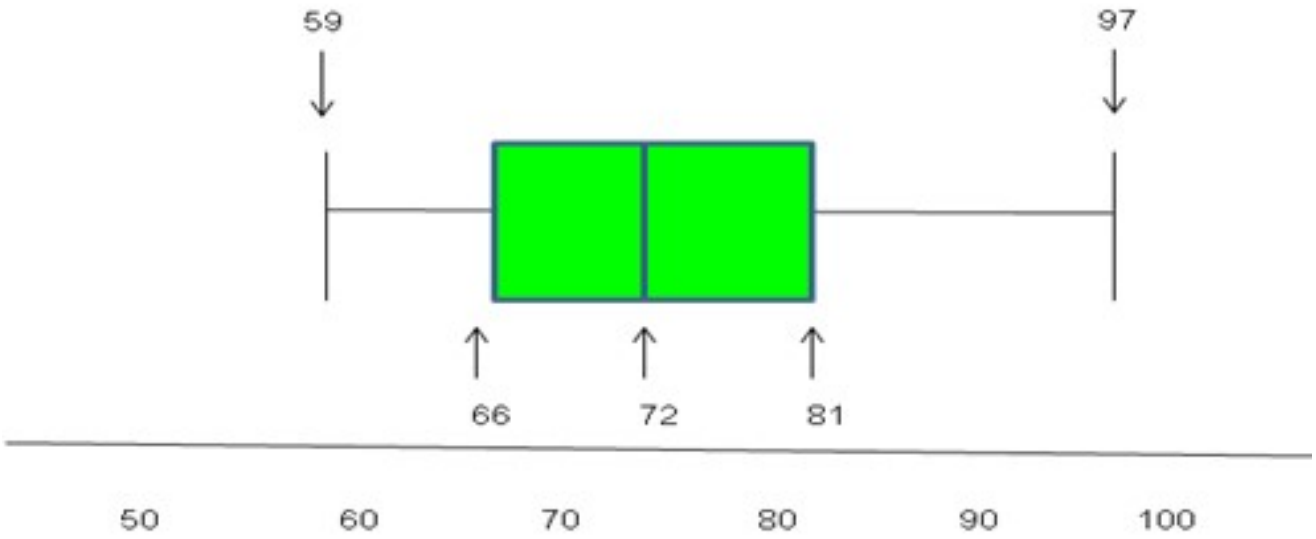
$$Q_3 \text{ อยู่ในตำแหน่งที่ } \frac{3(27+1)}{4} = 21 \text{ ดังนั้น } Q_3 = 81$$

ตัวอย่างที่ 5

3. หาค่า $Q_1 - 1.5(Q_3 - Q_1)$ ซึ่งคือ $66 - 1.5(81 - 66) = 43.5$ และหาค่า $Q_3 + 1.5(Q_3 - Q_1)$ ซึ่งคือ $81 + 1.5(81 - 66) = 103.5$

4. เนื่องจากไม่มีข้อมูลที่มีค่าน้อยกว่า 43.5 หรือมากกว่า 103.5 ดังนั้น ข้อมูลชุดนี้ไม่มีค่านอกเกณฑ์

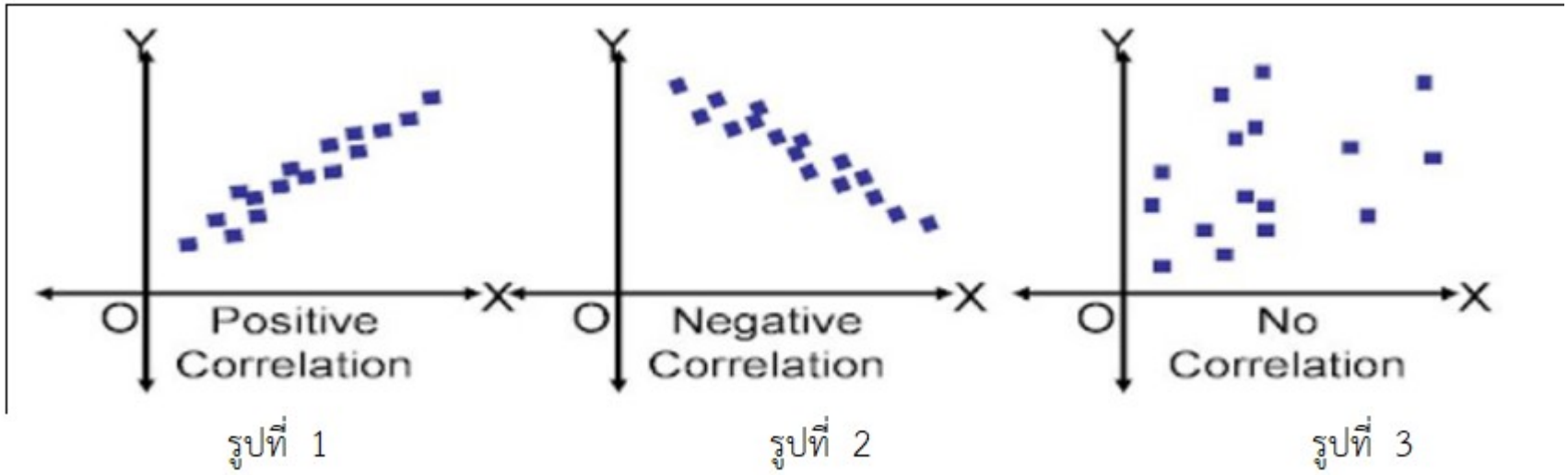
5. เขียนแผนภาพกล่องได้ดังนี้



แผนภาพการกระจาย

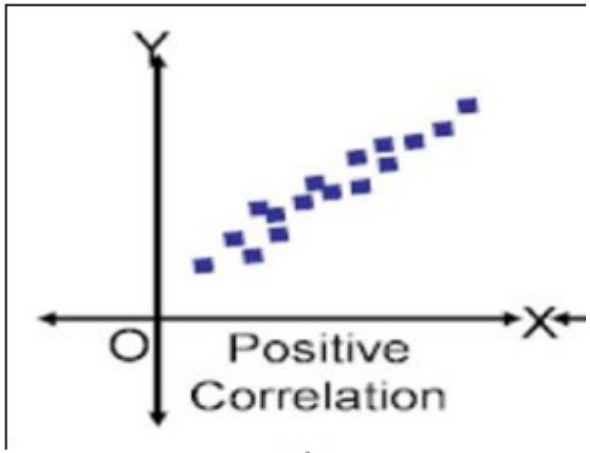
- แผนภาพการกระจาย (scatter plot) คือแผนภาพที่เกิดจากการลงจุดที่แสดงค่าของตัวแปรคู่หนึ่ง รูปแบบการกระจายของจุดต่าง ๆ ที่ปรากฏในแผรภาพจะแสดงถึงรูปแบบความสัมพันธ์ระหว่างสองตัวแปรนั้น

พิจารณาแผนภาพการกระจายต่อไปนี้



รูปที่ 1 รูปที่ 2 รูปที่ 3

แผนภาพการกระจาย

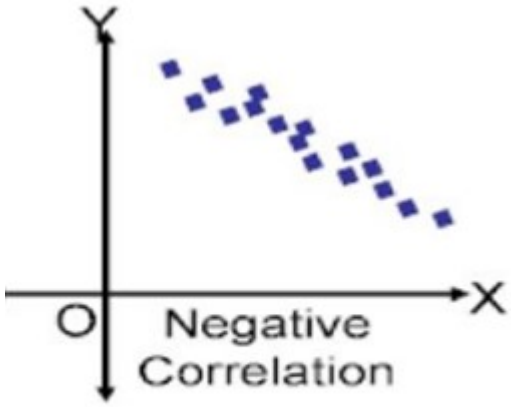


รูปที่ 1

- เมื่อกำหนดให้
  - รูปที่ 1 ค่าบนแกน X และ Y แทนพื้นที่ห้องชุดและราคาห้องชุดของคอนโดมิเนียมในย่านธุรกิจแห่งหนึ่งในกรุงเทพมหานคร ตามลำดับ

- จากรูปที่ 1 จะเห็นว่าเมื่อพื้นที่ห้องชุดมากขึ้น ราคาห้องชุดจะมากขึ้นด้วย ในกรณีนี้จะกล่าวว่าตัวแปรทั้งสองมีความสัมพันธ์ในทิศทางเดียวกัน

แผนภาพการกระจาย

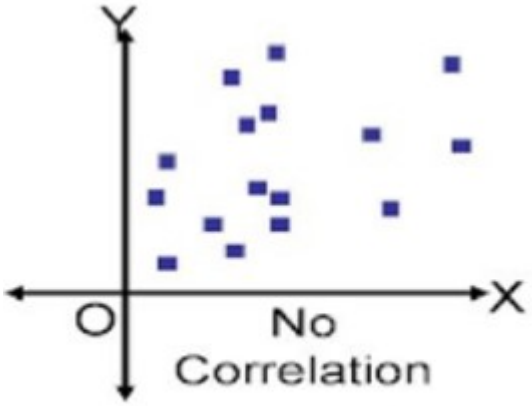


รูปที่ 2

- เมื่อกำหนดให้
  - รูปที่ 2 ค่าบนแกน X และ Y แทนอายุการใช้งานและมูลค่าของรถยนต์ยี่ห้อหนึ่งตามลำดับ

- จากรูปที่ 2 จะเห็นว่าเมื่ออายุการใช้งานมากขึ้น มูลค่าของรถยนต์จะยิ่งน้อยลง ในกรณีนี้จะหล่าวว่าตัวแปรทั้งสองมีความสัมพันธ์ในทิศทางตรงกันข้าม

แผนภาพการกระจาย



รูปที่ 3

- เมื่อกำหนดให้
  - รูปที่ 3 ค่าบนแกน X และ Y แทนอายุและรายได้ของประชากรวัยทำงานในหมู่บ้านตามลำดับ
  - จากรูปที่ 3 จะเห็นว่าเมื่ออายุมากขึ้น รายได้ของประชากรวัยทำงานในหมู่บ้านแห่งนี้ไม่ได้มากขึ้นหรือน้อยลงตาม ในกรณีนี้จะกล่าวว่าตัวแปรทั้งสองไม่มีความสัมพันธ์กัน

หมายเหตุ

ในที่นี้จะพิจารณาเฉพาะความสัมพันธ์เชิงเส้นเท่านั้น

แบบฝึกหัด 3.2

1. คะแนนสอบย่อยวิชาคณิตศาสตร์ซึ่งมีคะแนนเต็ม 10 คะแนน ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ห้องหนึ่ง จำนวน 30 คน แสดงได้ดังนี้

คะแนน	0	1	2	3	4	5	6	7	8	9	10
ความถี่	1	0	2	3	2	5	3	2	6	5	1

จงเขียนฮิสโทแกรมเพื่อนำเสนอข้อมูลชุดนี้

แบบฝึกหัด 3.2

2. โรงเรียนประจำจังหวัดแห่งหนึ่งในภาคใต้ได้จัดโครงการตรวจสุขภาพครู ประจำปี 2562 โดยผลการตรวจระดับน้ำตาลในเลือด (มิลลิกรัมต่อเดซิลิตร) ของครูทั้งหมด 80 คน แสดงได้ดังนี้

ระดับน้ำตาลในเลือด (มิลลิกรัมต่อเดซิลิตร)	ความถี่สะสม
61 – 67	3
68 – 74	6
75 – 81	9
82 – 88	24
89 – 95	47
96 – 102	63
103 – 109	76
110 - 116	86

แบบฝึกหัด 3.2

1) ครูมีระดับน้ำตาลในเลือดอยู่ในช่วงใดมากที่สุด

2) จงเขียนฮิสโทแกรมจากข้อมูลที่กำหนดให้

3) ถ้าระดับน้ำตาลในเลือดอยู่ในช่วง 100 - 125 มิลลิกรัมต่อเดซิลิตร จะถือว่ามีความเสี่ยงต่อการเป็นโรคเบาหวาน จงพิจารณาว่าข้อความ "มีครูมากกว่าร้อยละ 42 ของครูทั้งหมดมีความเสี่ยงต่อการเป็นโรคเบาหวาน" เป็นจริงหรือไม่ เพราะเหตุใด

3. จำนวนภาพยนต์ (เรื่อง) ที่นักเรียนชั้นมัธยมศึกษาปีที่ 6 จำนวน 50 คน ชมในหนึ่งปี เป็นดังนี้

33	17	15	18	20	10	15	5	22	25
22	6	16	24	25	17	7	20	19	3
12	16	21	32	16	16	23	25	28	25
26	27	9	17	5	20	15	16	7	19
32	17	16	7	18	26	28	18	16	10

1) จงเขียนแผนภาพจุดและแผนภาพลำต้นและใบจากข้อมูลที่กำหนดให้

2) นักเรียนที่ชมภาพยนตร์มากกว่า 12 เรื่องในหนึ่งปี คิดเป็นร้อยละเท่าใดของจำนวนนักเรียนทั้งหมด

แบบฝึกหัด 3.2

4. นภาพักตร์ได้จดบันทึกจำนวนรถจักรยานยนต์ที่มาจอดบริเวณหน้าบ้านในช่วงเวลา 08:00 – 09:00 น. ของแต่ละวัน เป็นเวลา 1 เดือน ได้ข้อมูลดังนี้

10	6	7	12	13	15	8	6	10	16
17	20	18	5	9	9	7	10	11	18
19	15	16	17	20	16	12	14	18	14
17									

1) จงหาควอร์ไทล์ที่ 1 ควอร์ไทล์ที่ 2 และควอร์ไทล์ที่ 3 ของข้อมูลชุดนี้

2) ข้อมูลชุดนี้มีค่นอกเกณฑ์หรือไม่ ถ้ามีคือค่าใด

3) จงเขียนแผนภาพกล่องเพื่อนำเสนอข้อมูลชุดนี้

4) จากแผนภาพกล่องที่ได้ในข้อ 3) จงอธิบายลักษณะการกระจายของข้อมูลชุดนี้

แบบฝึกหัด 3.2

5. ข้อมูลความสูง (เซนติเมตร) และน้ำหนัก (กิโลกรัม) ของนักเรียนที่สุ่มมาจำนวน 15 คน แสดงได้ดังนี้

ความสูง (ซม.)	168	152	155	149	145	154	146	160	162	152	163	157	164	141	145
น้ำหนัก (กก.)	60	40	45	41	34	33	40	42	55	39	60	46	50	30	39

จงเขียนแผนภาพการกระจายของข้อมูลชุดนี้ พร้อมทั้งพิจารณาว่าความสูงและน้ำหนักของนักเรียนมีความสัมพันธ์กันหรือไม่ อย่างไร

ค่าวัดทางสถิติ

- ประกอบด้วย
  - 1. ค่ากลางของข้อมูล
  - 2. ค่าวัดการกระจาย
  - 3. ค่าวัดตำแหน่งที่ของข้อมูล

ค่ากลางของข้อมูล : ค่าเฉลี่ยเลขคณิต

- **ค่าเฉลี่ยเลขคณิต (arithmetic mean)**เป็นค่าที่หาได้จากการหาผลรวมของข้อมูลทั้งหมดด้วยจำนวนข้อมูลที่มี

ให้ $x_1, x_2, x_3, \ldots, x_N$ แทนข้อมูล เมื่อ N แทนขนาดประชากร

ค่าเฉลี่ยเลขคณิตของประชากร (population mean) เขียนแทนด้วย μ (อ่านว่า มิว) หาได้จาก

$$\mu = \frac{x_1 + x_2 + x_3 + \cdots + x_N}{N}$$

ให้ $x_1, x_2, x_3, \ldots, x_n$ แทนข้อมูล เมื่อ n แทนขนาดตัวอย่าง

ค่าเฉลี่ยเลขคณิตของตัวอย่าง (sample mean) เขียนแทนด้วย $\bar{x}$ (อ่านว่า มิว) หาได้จาก

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \cdots + x_n}{n}$$

ค่ากลางของข้อมูล : ค่าเฉลี่ยเลขคณิต

เพื่อความสะดวกจะใช้ตัวอักษรกรีกตัวพิมพ์ใหญ่ Σ (อ่านว่า ซิกมา) เป็นสัญลักษณ์แสดงการบวก กล่าวคือ จะเขียนแทน $x_1 + x_2 + x_3 + \cdots + x_n$ ด้วยสัญลักษณ์ $\sum_{i-1}^{n} x_i$ (อ่านว่า ซัมเมชัน x_i เมื่อ i เท่ากับ 1 ถึง n) ดังนั้น สามารถเขียนสูตรของค่าเฉลี่ยเลขคณิตได้ดังนี้

$$\mu = \frac{\sum_{i-1}^{N} x_i}{N} \text{ และ } \bar{x} = \frac{\sum_{i-1}^{n} x_i}{n}$$

หมายเหตุ

การหาค่าเฉลี่ยเลขคณิตจะใช้ได้กับข้อมูลเชิงปริมาณเท่านั้น เช่น สามารถหาค่าเฉลี่ยเลขคณิตของอายุ รายได้ ความสูง แต่จะไม่สามารถหาค่าเฉลี่ยเลขคณิตของข้อมูลเชิงคุณภาพ เช่น ยี่ห้อรถยนต์ เชื้อชาติ เบอร์โทรศัพท์

ตัวอย่างที่ 5

ทีมฟุตบอลชายของโรงเรียนแห่งหนึ่งมีสมาชิกทั้งหมด 24 คน โดยความสูง (เซนติเมตร) ของสมาชิกแต่ละคนแสดงได้ดังนี้

165	178	170	168	167	167	180	175	181	164
179	158	177	163	165	172	180	191	185	176
175	183	177	179						

จงหาค่าเฉลี่ยเลขคณิตของข้อมูลชุดนี้

ให้ x_i แทนความสูงของนักฟุตบอลคนที่ i เมื่อ $i \in \{1,2,3,\ldots,24\}$ เนื่องจากข้อมูลชุดนี้เป็นข้อมูลของประชากร จึงให้ μ แทนค่าเฉลี่ยเลขคณิตของข้อมูลชุดนี้ จะได้

$$\mu = \frac{\sum_{i-1}^{24} x_i}{24} = \frac{4175}{24} \approx 173.96$$

ดังนั้น ค่าเฉลี่ยเลขคณิตของข้อมูลชุดนี้มีค่าประมาณ 173.96 เซนติเมตร

ตัวอย่างที่ 6

โรงเรียนแห่งหนึ่งกำหนดว่านักเรียนจะได้เกรด 4 วิชาคณิตศาสตร์ ก็ต่อเมื่อนักเรียนได้คะแนนเฉลี่ยจากการสอบย่อย 6 ครั้ง ไม่ต่ำกว่า 80 คะแนน ถ้าค่าเฉลี่ยเลขคณิตจากการสอบย่อย 5 ครั้ง ของนักเรียนคนหนึ่งเท่ากับ 77 คะแนน จงหาว่าในการสอบย่อยครั้งที่ 6 นักเรียนคนนี้จะต้องได้คะแนนอย่างน้อยเท่าใด จึงจะได้เกรด 4

ให้ x_i แทนคะแนนสอบย่อยครั้งที่ i เมื่อ $i \in \{1,2,3,\ldots,6\}$ จากโจทย์จะได้ $\frac{x_1+x_2+x_3+x_4+x_5}{5} = 77$ ดังนั้น $x_1+x_2+x_3+x_4+x_5 = 385$

ถ้าต้องการได้เกรด 4 วิชาคณิตศาสตร์ จะต้องได้คะแนนเฉลี่ยจากการสอบย่อย 6 ครั้ง ไม่ต่ำกว่า 80 คะแนน นั่นคือ $\frac{x_1+x_2+x_3+x_4+x_5+x_6}{6} \geq 80$ จะได้ $\frac{385+x_6}{6} \geq 80$ นั่นคือ $x_6 \geq 95$

ดังนั้น ในการสอบย่อยครั้งที่ 6 นักเรียนคนนี้จะต้องได้คะแนนอย่างน้อย 95 คะแนน จึงจะได้เกรด 4 วิชาคณิตศาสตร์

ตัวอย่างที่ 7

คะแนนสอบวิชาคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ที่สุ่มตัวอย่างมาจากห้องหนึ่งจำนวน 11 คน เป็นดังนี้

70 72 68 3 71 74 70 67 73 5 78

จงหา

1) ค่านอกเกณฑ์ของข้อมูลชุดนี้

2) ค่าเฉลี่ยเลขคณิตของข้อมูลชุดนี้

3) ค่าเฉลี่ยเลขคณิตของคะแนนสอบวิชาคณิตศาสตร์ของนักเรียนที่สุ่มขึ้นมา โดยไม่รวมค่านอกเกณฑ์

ตัวอย่างที่ 7

1) เรียงข้อมูลจากน้อยไปมากได้ดังนี้

3 5 67 68 70 70 71 72 73 74 78

จากข้อมูลข้างต้นจะได้ว่า $Q_1 = 67$ และ $Q_3 = 73$ จะได้ว่า $Q_1 - 1.5(Q_3 - Q_1) = 58$ และ $Q_3 + 1.5(Q_3 - Q_1) = 82$ ซึ่งทำให้ได้ว่าค่านอกเกณฑ์คือ 3 และ 5

2) ค่าเฉลี่ยเลขคณิตของข้อมูล ประมาณ 59.18

3) ค่าเฉลี่ยเลขคณิตของคะแนนสอบวิชาคณิตศาสตร์ของนักเรียนที่สุ่มขึ้นมา โดยไม่รวมค่านอกเกณฑ์ ประมาณ 71.44

ค่ากลางของข้อมูล : ค่าเฉลี่ยเลขคณิตถ่วงน้ำหนัก

- **ค่าเฉลี่ยถ่วงน้ำหนัก (weighted arithmetic mean)** เหมาะสำหรับใช้ในกรณีที่ข้อมูลแต่ละค่ามีความสำคัญไม่เท่ากัน

ให้ $x_1, x_2, x_3, \ldots, x_N$ แทนข้อมูล เมื่อ N แทนขนาดประชากร และให้ $w_1, w_2, w_3, \ldots, w_N$ แทนน้ำหนักของข้อมูล $x_1, x_2, x_3, \ldots, x_N$ ตามลำดับ จะได้ ค่าเฉลี่ยเลขคณิตถ่วงน้ำหนัก เท่ากับ $\mu = \frac{\sum_{i=1}^{N} w_i x_i}{\sum_{i=1}^{N} w_i}$

ตัวอย่างที่ 8

.ในการคำนวณเกรดเฉลี่ย (grade point average: GPA) ของนักเรียนคนหนึ่ง สมมติว่านักเรียนคนนี้ลงทะเบียนเรียน 5 วิชา ซึ่งแต่ละวิชามีหน่วยกิตไม่เท่ากัน และได้เกรดแต่ละวิชาดังนี้

วิชา	คณิตศาสตร์	ภาษาอังกฤษ	ภาษาไทย	สังคมศึกษา	สุขศึกษา
หน่วยกิต	2.0	3.0	3.0	3.0	1.0
เกรด	4	4	3	3	2

จงหาเกรดเฉลี่ยของนักเรียนคนนี้

วิชา	คณิตศาสตร์	ภาษาอังกฤษ	ภาษาไทย	สังคมศึกษา	สุขศึกษา	รวม
หน่วยกิต (w_i)	2.0	3.0	3.0	3.0	1.0	12
เกรด (x_i)	4	4	3	3	2	
$w_i x_i$	8	12	9	9	3	40

ดังนั้น เกรดเฉลี่ยของนักเรียนคนนี้ เท่ากับ 40/12 ประมาณ 3.33

ค่ากลางของข้อมูล : มัธยฐาน

- **มัธยฐาน (median)** คือ ค่าที่อยู่ในตำแหน่งกึ่งกลางของข้อมูลเมื่อเรียงข้อมูลจากน้อยไปมาก

ถ้าข้อมูลมี n ตัว การหามัธยฐานทำได้โดยเรียงลำดับข้อมูลจากน้อยไปมาก จะได้มัธยฐานอยู่ในตำแหน่งที่ $\frac{n+1}{2}$ นั่นคือ

- ถ้า n เป็นจำนวนคี่ มัธยฐานคือข้อมูลที่อยู่กึ่งกลาง
- ถ้า n เป็นจำนวนคู่ มัธยฐานคือค่าเฉลี่ยของข้อมูลสองตัวที่อยู่กึ่งกลาง

ตัวอย่างที่ 9

ความสูง (เซนติเมตร) ของนักเรียนหญิงจำนวน 11 คน แสดงได้ดังนี้

164 158 167 160 163 159 162 161 155 170 160

จงหามัธยฐานของข้อมูลชุดนี้

เรียงความสูงของนักเรียนหญิง 11 คน จากน้อยไปมาก ได้ดังนี้

155 158 159 160 161 162 163 164 167 168 170

เนื่องจากมัธยฐานอยู่ในตำแหน่งที่ $\frac{11+1}{2} = 6$

ดังนั้น มัธยฐานของข้อมูลชุดนี้ คือ 162 เซนติเมตร

ตัวอย่างที่ 10

ระยะเวลา (นาที) ที่ใช้ในการเดินทางจากบ้านำปโรงเรียนของนักเรียน 6 คน แสดงได้ดังนี้

32 15 45 12 90 25

จงหามัธยฐานของข้อมูลชุดนี้

เรียงระยะเวลาที่ใช้ในการเดินทางจากบ้านไปโรงเรียนของนักเรียน 6 คน จากน้อยไปมาก ได้ดังนี้

12 15 25 32 45 90

เนื่องจากมัธยฐานอยู่ในตำแหน่งที่ $\frac{6+1}{2} = 3.5$

ดังนั้น มัธยฐานของข้อมูลชุดนี้ คือ ค่าเฉลี่ยเลขคณิตของข้อมูลในตำแหน่งที่ 3 และ 4 ซึ่งคือ $\frac{25+32}{2} =$ 28.5 นาที

ค่ากลางของข้อมูล : ฐานนิยม

- **ฐานนิยม (mode)** คือ ข้อมูลที่มีจำนวนครั้งของการเกิดซ้ำกันมากที่สุดหรือข้อมูลที่มีความถี่สูงสุดที่มากกว่า 1 ซึ่งในที่นี้จะพิจารณาเฉพาะข้อมูลที่มีฐานนิยมเพียงค่าเดียว

ตัวอย่างที่ 11

อายุ (ปี) ของนักเรียนที่มาเข้าค่ายคณิตศาสตร์ จำนวน 15 คน แสดงได้ดังนี้

8 7 6 7 8 12 11 10 11 8 6 8 7 8

จงหาฐานนิยมของข้อมูลชุดนี้

- ฐานนิยมของข้อมูลชุดนี้ คือ 8 ปี

ข้อสังเกตที่สำคัญเกี่ยวกับค่ากลางชนิดต่าง ๆ

- 1. ฐานนิยมจะมีค่าตรงกับค่าใดค่าหนึ่งของข้อมูลชุดนั้น ในขณะที่ค่าเฉลี่ยเลขคณิตและมัธยฐานอาจไม่ใช่ค่าใดค่าหนึ่งของข้อมูลชุดนั้น
- 2. โดยปกติค่าเฉลี่ยเลขคณิตมักเป็นค่ากลางที่นิยมมากที่สุด แต่ถ้าข้อมูลมีข้อมูลที่แตกต่างจากข้อมูลตัวอื่นมากจะมีผลต่อค่าเฉลี่ยเลขคณิตของข้อมูลชุดนี้ แต่จะไม่มีผลต่อมัธยฐานและฐานนิยม
- 3. สำหรับข้อมูลเชิงคุณภาพ จะสามารถหาค่ากลางได้เฉพาะฐานนิยมเท่านั้น ไม่สามารถหาค่าเฉลี่ยเลขคณิตและมัธยฐานได้
- 4. ค่าเฉลี่ยเลขคณิต มัธยฐาน และฐานนิยมไม่จำเป็นต้องมีค่าเท่ากัน ทั้งนี้ค่าเฉลี่ยเลขคณิตและมัธยฐานจะมีค่าที่ไม่สูงหรือจ่ำเกินไปเมื่อเทียบกับค่าของข้อมูลทั้งหมด ในขณะที่ฐานนิยมอาจเป็นค่าสูงสุดหรือค่าต่ำสุดของข้อมูลนั้นได้

ข้อสังเกตที่สำคัญเกี่ยวกับค่ากลางชนิดต่าง ๆ

- การพิจารณาเลือกใช้ค่ากลางของข้อมูลไม่มีกฎเกณฑ์ที่แน่ชัด แต่ควรเลือกให้เหมาะสมกับวัตถุประสงค์ในการนำไปใช้และลักษณะของข้อมูล รวมทั้งต้องพิจารณาข้อดีและข้อเสียของค่ากลางแต่ละชนิด หากเลือกใช้ค่ากลางที่ไม่เหมาะสมอาจทำให้สรุปผลหรือตัดสินใจผิดพลาดได้

ตัวอย่างที่ 12

เงินเดือน (บาท) ของพนักงานแผนกหนึ่งในบริษัทแห่งหนึ่งจำนวนทั้งหมด 7 คน แสดงได้ดังนี้

15300 16600 13450 15300 14400 15100 71000

จงหาค่าเฉลี่ยเลขคณิต มัธยฐาน และฐานนิยมของข้อมูลชุดนี้ และพิจารณาว่าควรใช้ค่ากลางใดเป็นตัวแทนของข้อมูลชุดนี้ พร้อมทั้งให้เหตุผลประกอบ

- ค่าเฉลี่ยเลขคณิตของข้อมูลชุดนี้ คือ 23050 บาท
- มัธยฐานของข้อมูลชุดนี้ 15300 บาท
- ฐานนิยมของข้อมูลชุดนี้ 15300 บาท
- สังเกตว่า 71000 เป็นค่าที่สูงกว่าค่าส่วนใหญ่ของข้อมูลชุดนี้ หากใช้ค่าเฉลี่ยเลขคณิตเป็นตัวแทนของข้อมูลชุดนี้ จะเห็นได้ว่ามีพนักงานจำนวน 6 จาก 7 คน ที่ได้รับเงินเดือนน้อยกว่าค่าเฉลี่ยเลขคณิต ดังนั้น ค่าเฉลี่ยเลขคณิตจึงไม่เหมาะสมที่จะเป็นตัวแทนของข้อมูลชุดนี้ อย่างไรก็ตาม 71000 ไม่มีผลต่อมัธยฐานและฐานนิยม และจากตัวอย่างนี้จะเห็นว่าทั้งสองค่านี้เท่ากัน
- ดังนั้น ควรใช้มัธยฐานและฐานนิยมเป็นตัวแทนของข้อมูลชุดนี้

ตัวอย่างที่ 13

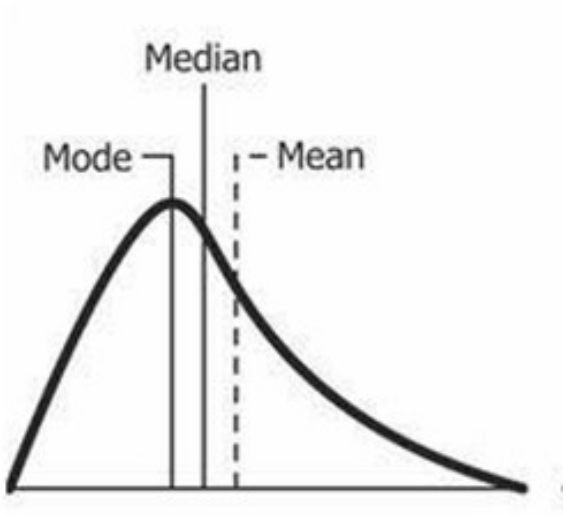
นักเรียนคนหนึ่งได้คะแนนสอบย่อยวิชาคณิตศาสตร์ จำนวน 5 ครั้ง ซึ่งแต่ละครั้งมีคะแนนเต็มเท่ากัน ดังนี้

17 17 17 19 20

จงพิจารณาว่าควรใช้ค่ากลางใดเป็นตัวแทนของข้อมูลชุดนี้ พร้อมทั้งให้เหตุผลประกอบ

- ค่าเฉลี่ยเลขคณิตของข้อมูลชุดนี้ คือ 18 คะแนน
- มัธยฐานของข้อมูลชุดนี้ 17 คะแนน
- ฐานนิยมของข้อมูลชุดนี้ 17 คะแนน
- จะเห็นว่าในสถานการณ์นี้ มัธยฐานและฐานนิยมไม่เหมาะสมที่จะเป็นตัวแทนของข้อมูลชุดนี้ เนื่องจากไม่ได้ใช้คะแนนสอบย่อยทุกครั้งในการคำนวณ ในขณะที่ค่าเฉลี่ยเลขคณิตได้จากการนำคะแนนสอบย่อยทุกครั้งมาคำนวณ และคะแนนสอบย่อยแต่ละครั้งใกล้เคียงกัน
- ดังนั้นควรใช้ค่าเฉลี่ยเลขคณิตเป็นตัวแทนของข้อมูลชุดนี้

ความสัมพันธ์ระหว่างการกระจายของข้อมูลและค่ากลางของข้อมูล

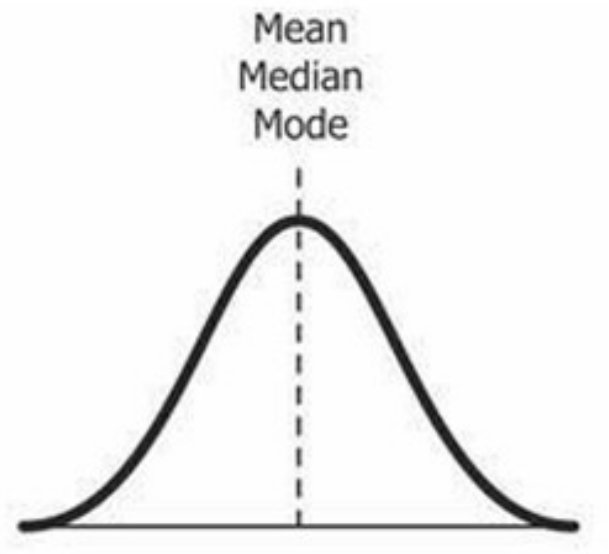


การแจกแจงเบ้ขวา

- การแจกแจงเบ้ขวา จะมีความสัมพันธ์ระหว่างค่ากลางคือ

ฐานนิยม < มัธยฐาน < ค่าเฉลี่ยเลขคณิต

ความสัมพันธ์ระหว่างการกระจายของข้อมูลและค่ากลางของข้อมูล

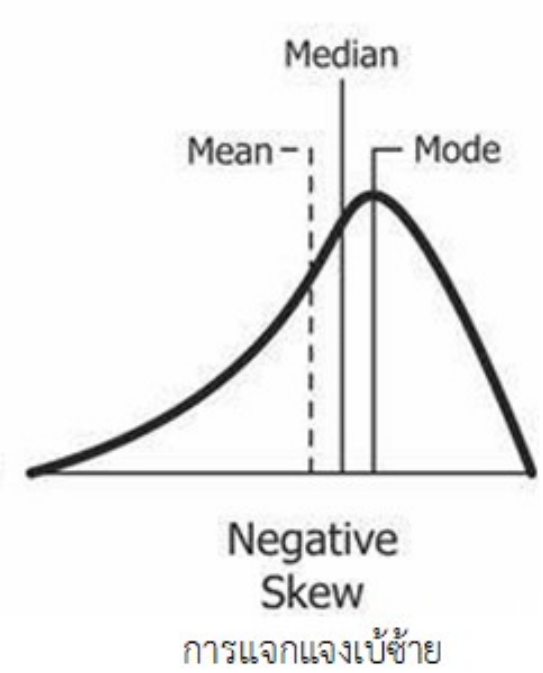


การแจกแจงสมมาตร

- การแจกแจงสมมาตร จะมีความสัมพันธ์ระหว่างค่ากลางคือ

ค่าเฉลี่ยเลขคณิต = มัธยฐาน = ฐานนิยม

ความสัมพันธ์ระหว่างการกระจายของข้อมูลและค่ากลางของข้อมูล



การแจกแจงเบ้ซ้าย

- การแจกแจงเบ้ซ้าย จะมีความสัมพันธ์ระหว่างค่ากลางคือ

ค่าเฉลี่ยเลขคณิต < มัธยฐาน < ฐานนิยม

แบบฝึกหัด 3.3

1. จากการสุ่มเก็บข้อมูลค่าจ้างรายวัน (บาท) ของพนักงานชั่วคราวของร้านสะดวกซื้อ 2 แห่ง เป็นเวลา 10 วัน ในเดือนมกราคม พ.ศ. 2562 แสดงได้ดังนี้

ร้านที่ 1	248	225	280	324	346	320	284	275	325	375
ร้านที่ 2	260	232	245	220	256	248	276	235	244	280

1) จงหาค่าเฉลี่ยเลขคณิต มัธยฐาน และฐานนิยมของค่าจ้างรายวันของพนักงานแต่ละร้าน

2) นักเรียนจะเลือกทำงานที่ร้านใด เพราะเหตุใด

2. ธนาคารแห่งหนึ่งสำรวจระยะเวลา (นาที) ที่ลูกค้าใช้ในการทำธุรกรรมที่ธนาคาร ได้ผลสำรวจดังนี้

14	13	17	15	15	14	15	28	18	17
11	9	13	16	18	15	14	16	7	16
11	12	19	27	14	12	19			

จงหา

แบบฝึกหัด 3.3

1) ค่านอกเกณฑ์ของข้อมูลชุดนี้

2) ค่าเฉลี่ยเลขคณิตของข้อมูลชุดนี้

3) ค่าเฉลี่ยเลขคณิตของระยะเวลาที่ลูกค้าใช้ในการทำธุรกรรมที่ธนาคารแห่งนี้ โดยไม่รวมค่านอกเกณฑ์

4) ค่าเฉลี่ยเลขคณิตที่ได้จากข้อ 2) และ 3) แตกต่างกันมากหรือไม่ เพราะเหตุใด

3. วิชาคณิตศาสตร์มีการสอบ 3 ครั้ง เป็นการสอบย่อย 2 ครั้ง และสอบปลายภาค 1 ครั้ง โดยคะแนนสอบย่อยแต่ละครั้งคิดเป็นร้อยละ 15 ของคะแนนทั้งหมด และคะแนนสอบปลายภาคคิดเป็นร้อยละ 70 ของคะแนนทั้งหมด ถ้านักเรียนคนหนึ่งได้คะแนนสอบย่อย 2 ครั้ง เป็น 74 และ 80 คะแนน และได้คะแนนสอบปลายภาค 62 คะแนน โดยแต่ละครั้งมีคะแนนเต็ม 100 คะแนน จงหาคะแนนเฉลี่ยวิชาคณิตศาสตร์ของนักเรียนคนนี้

แบบฝึกหัด 3.3

4. นักสัตววิทยาคนหนึ่งสำรวจจำนวนการตายและการเกิดของไก่ป่าชนิดหนึ่งในพื้นที่สำรวจ 14 พื้นที่ ในเดือนพฤษภาคม พ.ศ. 2562 ได้ผลสำรวจดังนี้

พื้นที่สำรวจ	จำนวนการตายของไก่ป่า (ตัว)	จำนวนการเกิดของไก่ป่า (ตัว)
1	5	30
2	0	28
3	0	38
4	8	34
5	9	26
6	7	40
7	2	48
8	6	46
9	4	32
10	0	31
11	2	46
12	10	132
13	3	42
14	5	126

แบบฝึกหัด 3.3

1) จงหาค่าเฉลี่ยเลขคณิต มัธยฐาน และฐานนิยมของจำนวนการตายของไก่ป่าในพื้นที่สำรวจทั้ง 14 พื้นที่ และพิจารณาว่าค่ากลางใดไม่ควรเป็นตัวแทนของจำนวนการตายของไก่ป่าในพื้นที่สำรวจทั้ง 14 พื้นที่ พร้อมทั้งให้เหตุผลประกอบ

2) จงหาค่าเฉลี่ยเลขคณิต มัธยฐาน และฐานนิยมของจำนวนการเกิดของไก่ป่าในพื้นที่สำรวจทั้ง 14 พื้นที่ และพิจารณาว่าควรใช้ค่ากลางใดเป็นตัวแทนของจำนวนการเกิดของไก่ป่าในพื้นที่สำรวจทั้ง 14 พื้นที่ พร้อมทั้งให้เหตุผลประกอบ

5. ในการสมัครเข้ารับการคัดเลือกเข้าศึกษาต่อในมหาวิทยาลัยแห่งหนึ่ง ผู้สมัครต้องยื่นจดหมายรับรองจำนวน 3 ฉบับ เพื่อประกอบการพิจารณา ถ้าค่าเฉลี่ยเลขคณิต มัธยฐาน และฐานนิยมของจำนวนจดหมายรับรองของผู้สมัครจำนวน 148 คน คือ 2.9, 3 และ 3 ฉบับ ตามลำดับ จงใช้ค่ากลางดังกล่าวในการอธิบายความหมายของจำนวนจดหมายรับรองของผู้สมัครทั้ง 148 คน

แบบฝึกหัด 3.3

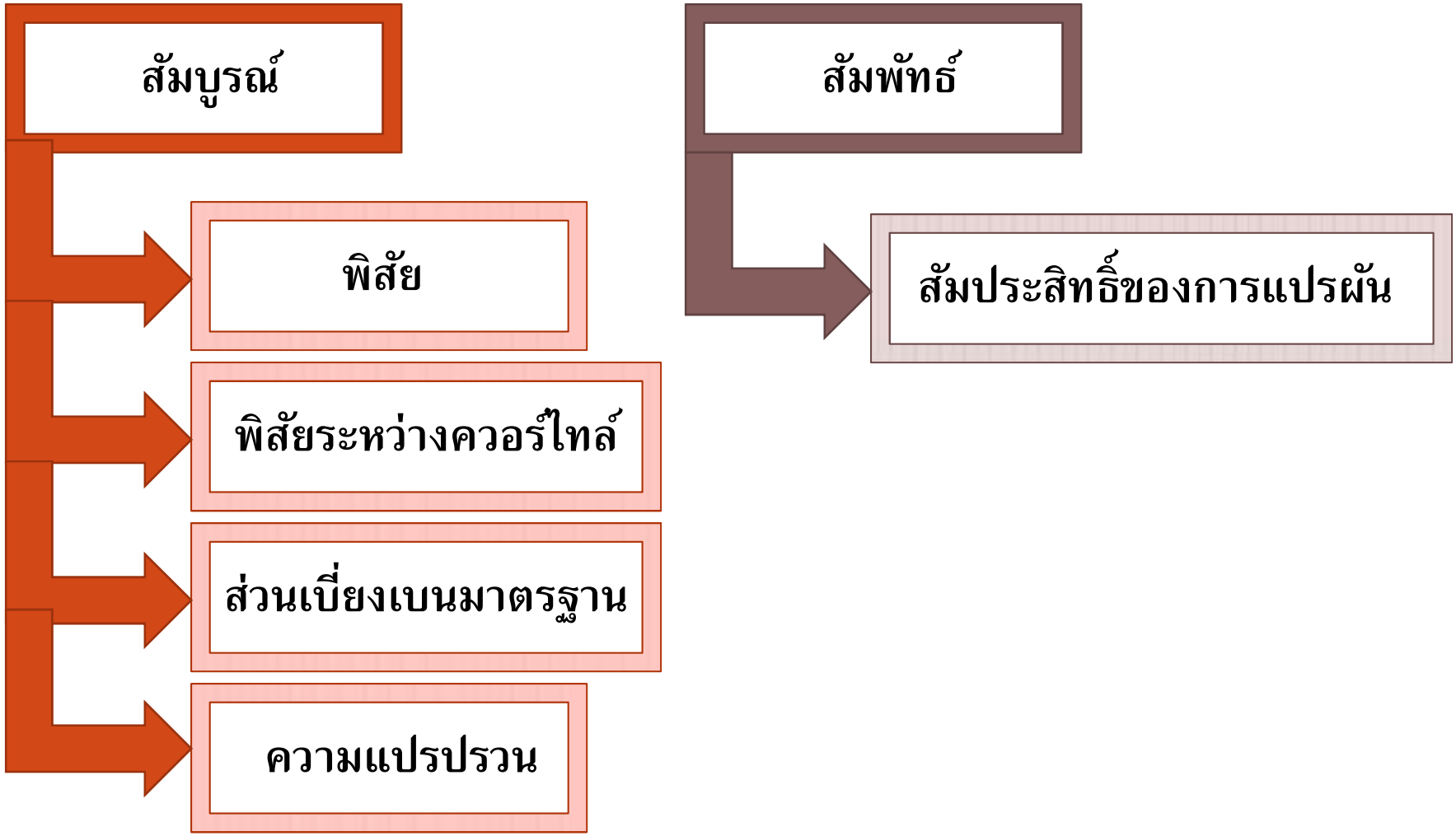
6. ค่าเฉลี่ยเลขคณิตของน้ำหนักของนักเรียน 3 คน คือ 38 กิโลกรัม และนักเรียนหนึ่งคนในกลุ่มนี้หนัก 46 กิโลกรัม ส่วนอีกสองคนที่เหลือหนักเท่ากัน จงหาว่านักเรียนสองคนที่เหลือหนักคนละกี่กิโลกรัม
7. ข้อมูลชุดหนึ่งมี 7 ตัว และมีค่าเฉลี่ยเลขคณิตคือ 81 ถ้าตัดข้อมูลออกไป 1 ตัว แล้วทำให้ค่าเฉลี่ยเลขคณิตของข้อมูลชุดนี้เหลือ 78 จงหาว่าข้อมูลที่ถูกตัดออกไปมีค่าเท่าใด
8. วิชาคณิตศสตร์มีการสอบย่อยทั้งหมด 5 ครั้ง แต่ละครั้งมีคะแนนเต็มเท่ากัน ถ้าค่าเฉลี่ยเลขคณิต มัธยฐาน และฐานนิยมของคะแนนสอบย่อยทั้งห้าครั้งของน้อยหน่า คือ 86, 87 และ 80 คะแนน ตามลำดับ จงหาคะแนนสอบย่อยที่สูงที่สุดที่เป็นไปได้ของน้อยหน่า ถ้าคะแนนสอบทั้งห้าครั้งของน้อยหน่าเป็นจำนวนเต็ม

ค่าวัดการกระจาย

- **การกระจายสัมบูรณ์ (absolute variation)** คือ การวัดการกระจายของข้อมูลด้วยค่าวัดทางสถิติที่มีหน่วยเช่นเดียวกับข้อมูลหรือเป็นกำลังสองของหน่วยของข้อมูล เพื่อใช้พิจารณาว่าข้อมูลแต่ละตัวมีความแตกต่างกันมากหรือน้อยเพียงใด

- **การกระจายสัมพัทธ์ (relative variation)** คือ การจัดการกระจายของข้อมูลด้วยค่าวัดทางสถิติที่ไม่มีหน่วย ซึ่งเป็นค่าที่ใช้ในการเปรียบเทียบการกระจายระหว่างข้อมูลมากกว่า 1 ชุด

ค่าวัดการกระจาย



การวัดการกระจายสัมบูรณ์ : พิสัย

- **พิสัย (range)** คือ ค่าที่ใช้วัดการกระจายของข้อมูลชุดหนึ่ง โดยคำนวณจากผลต่างระหว่างค่าสูงสุดและค่าต่ำสุดของข้อมูลชุดนั้น

พิสัย = ค่ามากที่สุด - ค่าน้อยที่สุด

ตัวอย่างที่ 14

ผลผลิตน้ำตาลใน พ.ศ. 2561/62 ของจีน สหรัฐอเมริกา ไทย อินเดีย ออสเตรเลีย และบราซิล แสดงได้ดังนี้

ประเทศ	จีน	สหรัฐอเมริกา	ไทย	อินเดีย	ออสเตรเลีย	บราซิล
ผลผลิต (ล้านตัน)	10.60	8.12	14.19	33.07	4.90	29.50

จงหาพิสัยของข้อมูลชุดนี้

- พิสัยของข้อมูลชุดนี้ คือ $33.07 - 4.90 = 28.17$ ล้านตัน

- จากตัวอย่างข้างต้น สามารถนำพิสัยมาใช้ในการอธิบายอย่างคร่าว ๆ ว่าผลผลิตน้ำตาลใน พ.ศ. 2561/62 ของแต่ละประเทศจะต่างกันไม่เกิน 28.17 ล้านตัน

การวัดการกระจายสัมบูรณ์ : พิสัยระหว่างควอร์ไทล์

- **พิสัยระหว่างควอร์ไทล์ (interquartile range)** คือ ค่าที่ใช้วัดการกระจายของข้อมูลโดยคำนวณจากผลต่างระหว่างควอร์ไทล์ที่สามและควอร์ไทล์ที่หนึ่ง เขียนแทนพิสัยระหว่างควอร์ไทล์ด้วย IQR

$$IQR = Q_3 - Q_1$$

- นอกจากนี้ IQR สามารถนำไปใช้ในการตรวจสอบว่าข้อมูลใดเป็นค่านอกเกณฑ์ โดยค่านอกเกณฑ์คือข้อมูลที่มีค่าน้อยกว่า $Q_1 - 1.5IQR$ หรือมากกว่า $Q_3 + 1.5IQR$

ตัวอย่างที่ 15

ปริมาณพลังงาน (กิโลแคลอรี) ของอาหารจานเดียว 7 รายการ ที่จำหน่ายในโรงอาหารของโรงเรียนแห่งหนึ่ง แสดงได้ดังนี้

อาหารจานเดียว	ปริมาณพลังงาน (กิโลแคลอรี)
ข้าวราดแกงเขียวหวานไก่	338
ข้าวราดแกงไตปลา	319
ข้าวราดแกงส้มผักรวม	255
ข้าวราดผัดเผ็ดหอยลาย	424
ข้าวราดแกงพะแนงหมู	40+
ข้าวราดแกงฉู่ฉี่ปลาทู	365
ข้าวราดผัดผักรวม	353

จงหาพิสัยระหว่างควอร์ไทล์ของข้อมูลชุดนี้

- จากข้อมูลข้างต้นจะได้ $Q_1 = 319$ และ $Q_3 = 409$ ดังนั้น $IQR = 409 - 319 = 90$ กิโลแคลอรี

การวัดการกระจายสัมบูรณ์ : ส่วนเบี่ยงเบนมาตรฐาน

- **ส่วนเบี่ยงเบนมาตรฐาน (standard deviation)** คือ ค่าที่ใช้วัดการกระจายของข้อมูลโดยเป็นค่าที่บอกให้ทราบว่าข้อมูลแต่ละตัวอยู่ห่างจากค่าเฉลี่ยเลขคณิตโดยเฉลี่ยประมาณเท่าใด

การวัดการกระจายสัมบูรณ์ : ส่วนเบี่ยงเบนมาตรฐาน

ให้ $x_1, x_2, x_3, \ldots, x_N$ แทนข้อมูล เมื่อ N แทนขนาดประชากร และให้ μ แทนค่าเฉลี่ยเลขคณิตของข้อมูลชุดนี้

ส่วนเบี่ยงเบนมาตรฐานของประชากร เขียนแทนด้วย σ (อ่านว่า ซิกมา) หาได้จาก

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

ให้ $x_1, x_2, x_3, \ldots, x_n$ แทนข้อมูล เมื่อ n แทนขนาดตัวอย่าง และให้ $\bar{x}$ แทนค่าเฉลี่ยเลขคณิตของข้อมูลชุดนี้

ส่วนเบี่ยงเบนมาตรฐานของตัวอย่าง เขียนแทนด้วย s หาได้จาก

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

ตัวอย่างที่ 16

ความสูง (เซนติเมตร) ของนักวอลเลย์บอลหญิงของโรงเรียนแห่งหนึ่งจำนวนทั้งหมด 10 คน แสดงได้ดังนี้

174 171 170 184 180 179 169 178 181 160

จงหาส่วนเบี่ยงเบนมาตรฐานของข้อมูลชุดนี้

จากข้อมูลข้างต้นจะได้

x_i	$x_i - \mu$	$(x_i - \mu)^2$
174		
171		
170		
184		
180		
179		
169		
178		
181		
160		

- จากสูตร

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

- หาค่าเฉลี่ยเลขคณิตโดย

$$\mu = \frac{1746}{10} = 174.6$$

x_i	$x_i - \mu$	$(x_i - \mu)^2$
174	-0.6	0.36
171	-3.6	12.96
170	-4.6	21.16
184	9.4	88.36
180	5.4	29.16
179	4.4	19.36
169	-5.6	31.36
178	3.4	11.56
181	6.4	40.96
160	-14.6	213.16
		$\sum_{i=1}^{10}(x_i - \mu)^2$=

x_i	$x_i - \mu$	$(x_i - \mu)^2$
174	-0.6	0.36
171	-3.6	12.96
170	-4.6	21.16
184	9.4	88.36
180	5.4	29.16
179	4.4	19.36
169	-5.6	31.36
178	3.4	11.56
181	6.4	40.96
160	-14.6	213.16

| | | $\sum_{i=1}^{10}(x_i - \mu)^2$= 468.40 |

- ดังนั้น $\sigma = \sqrt{\frac{468.40}{10}} \approx 6.84$
- นั่นคือ ส่วนเบี่ยงเบนมาตรฐานของข้อมูลชุดนี้มีค่าประมาณ 6.84 เซนติเมตร

การวัดการกระจายสัมบูรณ์ : ความแปรปรวน

- **ความแปรปรวน (variance)** คือ ค่าที่ใช้วัดการกระจายของข้อมูล โดยคำนวณจากกำลังสองของส่วนเบี่ยงเบนมาตรฐาน

การวัดการกระจายสัมบูรณ์ : ความแปรปรวน

ให้ $x_1, x_2, x_3, \ldots, x_N$ แทนข้อมูล เมื่อ N แทนขนาดประชากร และให้ μ แทนค่าเฉลี่ยเลขคณิตของข้อมูลชุดนี้

ความแปรปรวนของประชากร เขียนแทนด้วย σ^2 หาได้จาก

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

ให้ $x_1, x_2, x_3, \ldots, x_n$ แทนข้อมูล เมื่อ n แทนขนาดตัวอย่าง และให้ $\bar{x}$ แทนค่าเฉลี่ยเลขคณิตของข้อมูลชุดนี้

ความแปรปรวนของตัวอย่าง เขียนแทนด้วย s^2 หาได้จาก

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

ตัวอย่างที่ 17

ในการศึกษาอายุขัยเฉลี่ย (ปี) ของสัตว์เลี้ยงลูกด้วยน้ำนม นักวิทยาศาสตร์ได้สุ่มตัวอย่างสัตว์เลี้ยงลูกด้วยน้ำนมมา 10 ชนิด พบว่าอายุขัยเฉลี่ยของสัตว์แต่ละชนิดเป็นดังนี้

สัตว์เลี้ยงลูกด้วยน้ำนม	อายุขัยเฉลี่ย (ปี)
แมว	13
วัว	15
สุนัข	12
ลา	12
แพะ	8
หนูตะเภา	4
ม้า	20
หนู	10
กระต่าย	5
แกะ	12

จงหาส่วนเบี่ยงเบนมาตรฐานและความแปรปรวนของอายุขัยเฉลี่ยของสัตว์เลี้ยงลูกด้วยน้ำนม 10 ชนิดนี้

ให้ x_i แทนอายุขัยเฉลี่ยของสัตว์เลี้ยงลูกด้วยน้ำนมชนิดที่ i เมื่อ $i \in \{1,2,3,\ldots,10\}$ และ $\bar{x}$ แทนค่าเฉลี่ยเลขคณิตของอายุขัยเฉลี่ยของสัตว์เลี้ยงลูกด้วยน้ำนม 10 ชนิดนี้ จะได้ $\bar{x} = \frac{\sum_{i=1}^{10} x_i}{10} = \frac{110}{10} = 11$

ดังนั้น ค่าเฉลี่ยเลขคณิตของอายุขัยเฉลี่ยของสัตว์เลี้ยงลูกด้วยน้ำนม 10 ชนิดนี้ คือ 11 ปี จากข้อมูลข้างต้นจะได้

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
12		
15		
12		
12		
8		
4		
20		
10		
5		
12		
		$\sum_{i=1}^{10}(x_i - \bar{x})^2$=

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
12	1	1
15	4	16
12	1	1
12	1	1
8	-3	9
4	-7	49
20	9	81
10	-1	1
5	6	36
12	1	1
		$\sum_{i=1}^{10}(x_i - \bar{x})^2$= 196

- ดังนั้น $s = \sqrt{\frac{196}{10-1}} = \frac{14}{3} \approx 4.67$ และ $s^2 = \frac{196}{9} \approx 21.78$

- นั่นคือ ส่วนเบี่ยงเบนมาตรฐานของอายุขัยเฉลี่ยของสัตว์เลี้ยงลูกด้วยด้วยน้ำนม 10 ชนิดนี้มีค่าประมาณ 4.67 ปี และความแปรปรวนของอายุขัยเฉลี่ยของสัตว์เลี้ยงลูกด้วยน้ำนม 10 ชนิดนี้ มีค่าประมาณ 21.78 ปี

การวัดการกระจายสัมพัทธ์ : สัมประสิทธิ์ของการแปรผัน

สัมประสิทธิ์การแปรผันของประชากร $= \frac{\sigma}{|\mu|}$ เมื่อ $\mu \neq 0$

สัมประสิทธิ์การแปรผันของตัวอย่าง $= \frac{s}{|\bar{x}|}$ เมื่อ $\bar{x} \neq 0$

สัมประสิทธิ์การแปรผันของประชากร $= \frac{\sigma}{|\mu|} \times 100\%$ เมื่อ $\mu \neq 0$

สัมประสิทธิ์การแปรผันของตัวอย่าง $= \frac{s}{|\bar{x}|} \times 100\%$ เมื่อ $\bar{x} \neq 0$

- การเปรียบเทียบการกระจายของข้อมูลโดยใช้สัมประสิทธิ์การแปรผันนั้น ถ้าสัมประสิทธิ์การแปรผันของข้อมูลชุดใดมีค่ามากกว่า หมายความว่าข้อมูลชุดนั้นมีการกระจายออกจากค่าเฉลี่ยเลขคณิตมากกว่า หรือกล่าวได้ว่าข้อมูลชุดนั้นเกาะกลุ่มกันน้อยกว่าข้อมูลอีกชุดหนึ่ง

ตัวอย่างที่ 18

ในการเปรียบเทียบคะแนนสอบวิชาคณิตศาสตร์และวิชาภาษาอังกฤษของนักเรียนห้องหนึ่งซึ่งมีคะแนนเต็มวิชาละ 100 คะแนน ครูประจำชั้นได้สุ่มตัวอย่างนักเรียนห้องนี้มา 10 คน พบว่าคะแนนสอบแต่ละวิชาของนักเรียนแต่ละคนเป็นดังนี้

นักเรียนคนที่	1	2	3	4	5	6	7	8	9	10
คะแนนสอบวิชาคณิตศาสตร์	56	62	76	90	78	81	88	79	80	75
คะแนนสอบวิชาภาษาอังกฤษ	78	74	63	89	76	75	85	90	73	74

จงหาสัมประสิทธิ์การแปรผันของคะแนนสอบวิชาคณิตศาสตร์และวิชาภาษาอังกฤษของนักเรียนที่สุ่มตัวอย่างมา 10 คนนี้ พร้อมทั้งเปรียบเทียบการกระจายของคะแนนสอบทั้งสองวิชาของนักเรียนที่สุ่มตัวอย่างมา 10 คนนี้

ให้ x_i และ y_i แทนคะแนนสอบวิชาคณิตศาสตร์และวิชาภาษาอังกฤษของนักเรียนคนที่ i เมื่อ $i \in \{1,2,3,\ldots,10\}$ ตามลำดับ

$\bar{x}$ และ $\bar{y}$ แทนค่าเฉลี่ยเลขคณิตของคะแนนสอบวิชาคณิตศาสตร์และวิชาภาษาอังกฤษที่สุ่มตัวอย่างมา 10 คนนี้ ตามลำดับ

s_x และ s_y แทนส่วนเบี่ยงเบนมาตรฐานของคะแนนสอบวิชาคณิตศาสตร์และวิชาภาษาอังกฤษของนักเรียนที่สุ่มตัวอย่างมา 10 คนนี้ ตามลำดับ

จะได้ $\bar{x} = \frac{\sum_{i=1}^{10} x_i}{10} = 76.7$ และ $\bar{y} = \frac{\sum_{i=1}^{10} y_i}{10} = 77.7$

x_i	$x_i - \overline{x}$	$(x_i - \overline{x})^2$	y_i	$y_i - \overline{y}$	$(y_i - \overline{y})^2$
58			78		
62			74		
76			63		
90			89		
78			76		
81			75		
88			85		
79			90		
80			73		
75			74		
		$\sum_{i=1}^{10}(x_i - \bar{x})^2=$			$\sum_{i=1}^{10}(x_i - \bar{x})^2=$

x_i	$x_i - \overline{x}$	$(x_i - \overline{x})^2$	y_i	$y_i - \overline{y}$	$(y_i - \overline{y})^2$
58	-18.7	349369	78	0.3	0.09
62	-14.7	216.09	74	-3.7	13.69
76	-0.7	0.49	63	-14.7	216.09
90	13.3	176.89	89	11.3	127.09
78	1.3	1.69	76	-1.7	2.89
81	4.3	18.49	75	-2.7	7.29
88	11.3	127.69	85	7.3	53.29
79	2.3	5.29	90	12.3	151.29
80	3.3	10.89	73	-4.7	22.04
75	-1.7	2.89	74	-3.7	13.69
		$\sum_{i=1}^{10}(x_i - \overline{x})^2 = 910.1$			$\sum_{i=1}^{10}(x_i - \overline{x})^2 = 608.1$

ดังนั้น $s_x = \sqrt{\frac{910.1}{10-1}} \approx 10.06$ และ $s_y = \sqrt{\frac{608.1}{10-1}} \approx 8.22$

นั่นคือ สัมประสิทธิ์การแปรผันของคะแนนสอบวิชาคณิตศาสตร์ของนักเรียนที่สุ่มตัวอย่างมา 10 คนนี้ คือ $\frac{10.06}{|76.7|} \approx 0.1312$ และสัมประสิทธิ์การแปรผันของคะแนนสอบวิชาภาษาอังกฤษของนักเรียนที่สุ่มตัวอย่างมา 10 คนนี้ คือ $\frac{8.22}{|77.7|} \approx 0.1058$

- เมื่อพิจารณาจากนักเรียนที่สุ่มตัวอย่างมา 10 คนนี้ จะเห็นว่าสัมประสิทธิ์การแปรผันของคะแนนสอบวิชาคณิตศาสตร์มากกว่าสัมประสิทธิ์การแปรผันของคะแนนสอบวิชาภาษาอังกฤษ สรุปได้ว่า คะแนนสอบวิชาคณิตศาสตร์มีการกระจายมากกว่าคะแนนสอบวิชาภาษาอังกฤษ หรือกล่าวได้ว่าคะแนนสอบวิชาภาษาอังกฤษเกาะกลุ่มกันมากกว่าคะแนนสอบวิชาคณิตศาสตร์

ข้อสังเกต สัมประสิทธิ์การแปรผันไม่มีหน่วย

แบบฝึกหัด 3.4

1. ร้านค้าจำหน่ายและรับติดตั้งประตูอัตโนมัติแห่งหนึ่งเก็บข้อมูลตัวอย่างเกี่ยวกับเวลา (นาที) ที่ใช้ในการติดตั้งประตูแต่ละบาน ได้ข้อมูลดังนี้

28 32 24 46 44 40 54 38 32 42 36

จงหาพิสัย พิสัยระหว่างควอร์ไทล์ ส่วนเบี่ยงเบนมาตรฐาน และความแปรปรวนของเวลาที่ใช้ในการติดตั้งประตู

2. ปริมาณพลังงาน (กิโลแคลอรี) ของอาหารจานเดียว 11 รายการ ที่สุ่มตัวอย่างมาจากโรงอาหารแห่งหนึ่ง แสดงได้ดังนี้

อาหารจานเดียว	ปริมาณพลังงาน (กิโลแคลอรี)
หอยทอด	933
สุกี้น้ำรวมมิตร	117
ข้าวผัดหมู	553
ข้าวหมูแดง	444
ข้าวมันไก่	717
เส้นใหญ่ราดหน้าหมู	337
ข้าวหมกไก่	475
ข้าวคลุกกะปิ	522
หมี่กรอบราดหน้าทะเล	344
ผัดไทยกุ้งสด	519
ขนมจีนแกงเขียวหวาน	337

1) จงหาพิสัยและพิสัยระหว่างควอร์ไทล์ของปริมาณพลังงานของอาหารจานเดียวที่สุ่มตัวอย่างมา 11 รายการนี้

2) จงพิจารณาว่าระหว่างพิสัยและพิสัยระหว่างควอร์ไทล์ ค่าวัดการกระจายสัมบูรณ์ใดเหมาะสำหรับใช้อธิบายลักษณะการกระจายของข้อมูลชุดนี้ พร้อมทั้งให้เหตุผลประกอบ

3. ครอบครัวหนึ่งประกอบด้วยพ่อ แม่ และลูกอีก 3 คน มีอายุ 45, 42, 20, 17 และ 16 ปี ตามลำดับ จงหาส่วนเบี่ยงเบนมาตรฐานและความแปรปรวนของอายุของสมาชิกในครอบครัวนี้และจงหาว่าในอีก 5 ปีข้างหน้า ส่วนเบี่ยงเบนมาตรฐานของอายุสมาชิกในครอบครัวนี้จะเป็นอย่างไร

4. จากการรายงานของศูนย์ข้อมูลอุบัติเหตุ เพื่อเสริมสร้างวัฒนธรรมความปลอดภัยทางถนน พบว่า จำนวนผู้บาดเจ็บรวม (ราย) ตั้งแต่ 2556 - 2558 ในแต่ละวันของช่วง 7 วันอันตราย ของเทศกาลปีใหม่ แสดงได้ดังนี้

วันที่ 1	วันที่ 2	วันที่ 3	วันที่ 4	วันที่ 5	วันที่ 6	วันที่ 7
1236	1633	1664	1458	1506	1423	870

จงหาพิสัย พิสัยระหว่างควอร์ไทล์ และส่วนเบี่ยงเบนมาตรฐานของข้อมูลชุดนี้

5. ค่าเฉลี่ยเลขคณิตและส่วนเบี่ยงเบนมาตรฐานของคะแนนสอบวิชาคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 จำนวน 2 ห้องเรียน ซึ่งมีคะแนนเต็ม 100 คะแนน เป็นดังนี้

	ค่าเฉลี่ยเลขคณิค	ส่วนเบี่ยงเบนมาตรฐาน
ห้อง 1	73.2	4.8
ห้อง 2	52.4	3.6

จงเปรียบเทียบการกระจายของคะแนนสอบวิชาคณิตศาสตร์ของนักเรียนสองห้องนี้

แบบฝึกหัด 3.4

6. อุณหภูมิสูงสุดและอุณหภูมิต่ำสุด (องศาเซลเซียส) ของจังหวัดขอนแก่น ตั้งแต่ พ.ศ. 2549 - 2538 แสดงได้ดังนี้

พ.ศ.	2549	2550	2551	2552	2553	2554	2555	2556	2557	2558
อุณหภูมิสูงสุด	39.1	41.1	38.5	39.6	41.2	39.3	39.0	41.8	40.5	41.0
อุณหภูมิต่ำสุด	12.0	12.6	11.9	10.2	13.5	11.6	15.0	11.6	10.2	11.6

จงหาสัมประสิทธิ์การแปรผันของอุณหภูมิสูงสุดและอุณหภูมิต่ำสุดของจังหวัดขอนแก่นตั้งแต่ พ.ศ. 2549 - 2558 พร้อมทั้งเปรียบเทียบการกระจายของข้อมูลทั้งสองชุดนี้

ค่าวัดตำแหน่งที่ของข้อมูล

- ควอร์ไทล์ที่ r คือ ค่าที่มีจำนวนข้อมูลที่มีค่าน้อยกว่าค่านี้อยู่ประมาณ r ในสี่ของจำนวนข้อมูลทั้งหมด เขียนแทนด้วย Q_r

- เปอร์เซ็นไทล์ที่ r คือ ค่าที่มีจำนวนข้อมูลที่มีค่าน้อยกว่าค่านี้อยู่ประมาณ r ในหนึ่งร้อยของจำนวนข้อมูลทั้งหมด เขียนแทนด้วย P_r

ค่าวัดตำแหน่งที่ของข้อมูล

ขั้นตอนการหาควอร์ไทล์ที่ r หรือ เปอร์เซ็นไทล์ที่ r

- เรียงข้อมูลจากน้อยไปมาก
- ถ้า n เป็นจำนวนข้อมูลทั้งหมด สามารถหาตำแหน่งได้ดังนี้

Q_r อยู่ในตำแหน่งที่ $\frac{r(n+1)}{4}$

P_r อยู่ในตำแหน่งที่ $\frac{r(n+1)}{100}$

- ในกรณีที่คำนวณตำแหน่งแล้วมีทศนิยม ให้เทียบบัญญัติไตรยางศ์เพื่อหาค่าของข้อมูล

ข้อสังเกต

- ตำแหน่ง P_{25}

$$\frac{25}{100}(N+1) = \frac{1}{4}(N+1)$$

- ตำแหน่ง Q_1

- ตำแหน่ง P_{50}

$$\frac{50}{100}(N+1) = \frac{2}{4}(N+1)$$

- ตำแหน่ง Q_2

- ตำแหน่ง P_{75}

$$\frac{75}{100}(N+1) = \frac{3}{4}(N+1)$$

- ตำแหน่ง Q_3

ตัวอย่างที่ 19

ข้อมูลปริมาณการส่งออกข้าวไทยโดยประมาณ (พันตันข้าวสาร) ใน พ.ศ. 2560 จำแนกตามชนิดของข้าว จากศูนย์เทคโนโลยีสารสนเทศและการสื่อสาร สำนักงานปลัดกระทรวงพาณิชย์ โดยความร่วมมือจากกรมศุลกากร แสดงได้ดังตาราง

ชนิดของข้าว	ปริมาณการส่งออกโดยประมาณ (พันตันข้าวสาร)
ต้นข้าวขาว	4662
ปลายข้าวขาว	408
ต้นข้าวหอมมะลิ	1630
ปลายข้าวหอมมะลิ	669
ข้าวนึ่ง	3370
ข้าวเหนียว	214
ปลายข้าวเหนียว	303
ข้าวหอมไทย	213

จงหา

1) ควอร์ไทล์ที่ 1 ควอร์ไทล์ที่ 2 และควอร์ไทล์ที่ 3 ของปริมาณการส่งออกโดยประมาณของข้าว 8 ชนิด ใน พ.ศ. 2560

2) ชนิดของข้าวที่มีปริมาณการส่งออกน้อยกว่าควอร์ไทล์ที่ 1

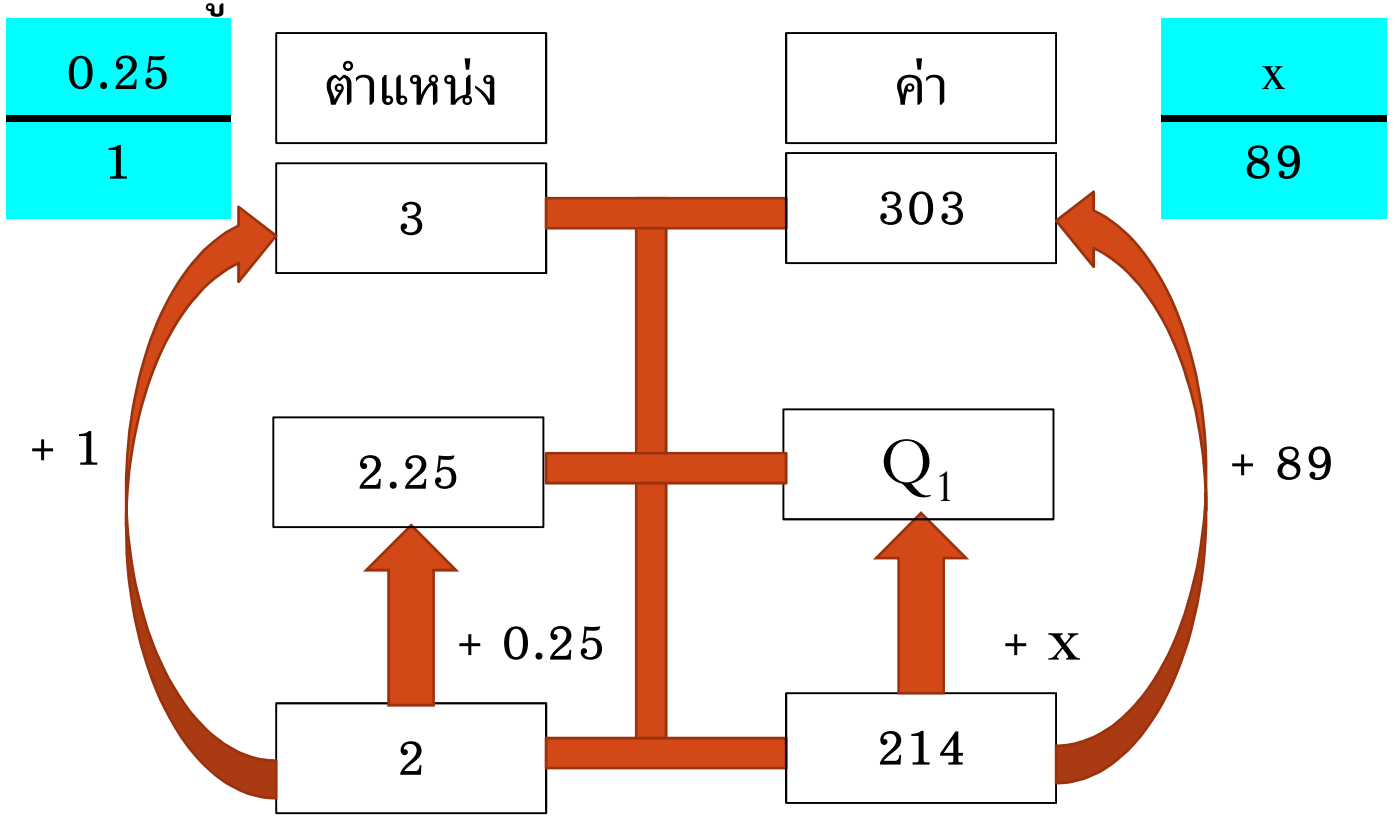
3) ชนิดของข้าวที่มีปริมาณการส่งออกมากกว่าควอร์ไทลที่ 3

เรียงปริมาณการส่งออกโดยประมาณของข้าว 8 ชนิด จากน้อยไปมาก ได้ดังนี้

213 214 303 408 669 1630 3370 4662

1) เนื่องจาก Q_1 อยู่ในตำแหน่งที่ $\frac{1}{4}(8+1) = 2.25$

ดังนั้น Q_1 อยู่ระหว่างข้อมูลในตำแหน่งที่ 2 และตำแหน่งที่ 3 ซึ่งมีค่าอยู่ระหว่าง 214 และ 303



$$\frac{0.25}{1} = \frac{x}{89}$$

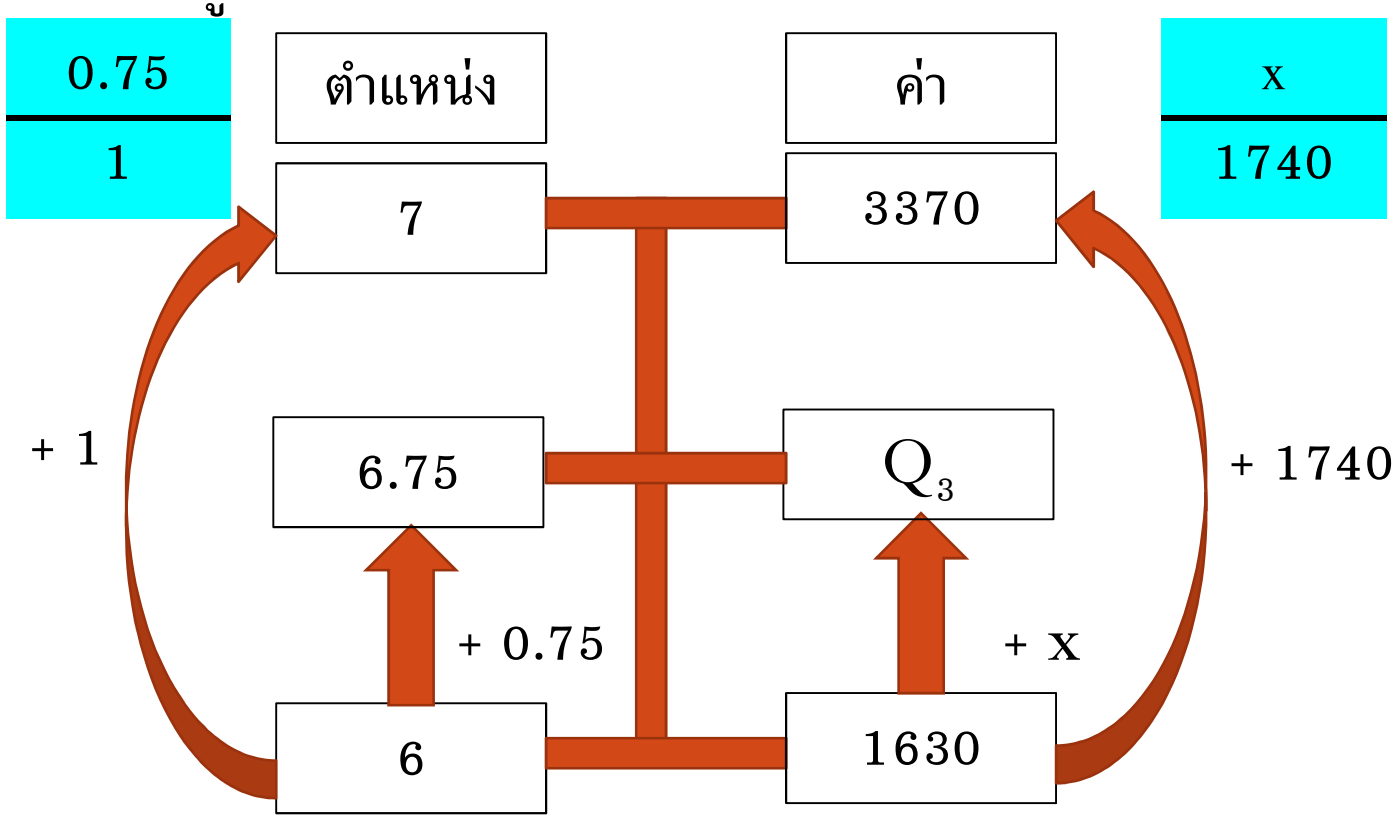
ดังนั้น $x = 0.25 \times 89 = 22.25$ นั่นคือ $Q_1 = 214 + 22.25 = 236.25$

เนื่องจาก Q_2 อยู่ในตำแหน่งที่ $\frac{2}{4}(8+1) = 4.5$

ดังนั้น Q_2 คือ ค่าเฉลี่ยเลขคณิตของข้อมูลในตำแหน่งที่ 4 และตำแหน่งที่ 5 ซึ่งคือ $\frac{408+669}{2} = 538.5$

เนื่องจาก Q_3 อยู่ในตำแหน่งที่ $\frac{3}{4}(8+1) = 6.75$

ดังนั้น Q_3 อยู่ระหว่างข้อมูลในตำแหน่งที่ 6 และตำแหน่งที่ 7 ซึ่งมีค่าอยู่ระหว่าง 1630 และ 3370



$$\frac{0.75}{1} = \frac{x}{1740}$$

ดังนั้น $x = 0.75 \times 1740 = 1305$ นั่นคือ $Q_3 = 1630 + 1305 = 2935$

จะได้ว่าควอร์ไทล์ที่ 1 ควอร์ไทล์ที่ 2 และควอร์ไทล์ที่ 3 ของปริมาณการส่งออกโดยประมาณของข้าว 8 ชนิด ใน พ.ศ. 2560 คือ 236.25, 538.5 และ 2935 พันตันข้าวสาร ตามลำดับ

2) ชนิดของข้าวที่มีปริมาณการส่งออกน้อยกว่าควอร์ไทล์ที่ 1 ซึ่งเท่ากับ 236.25 พันตันข้าวสาร ได้แก่ ข้าวหอมไทย และข้าวเหนียว

3) ชนิดของข้าวที่มีปริมาณการส่งออกมากกว่าควอร์ไทล์ที่ 3 ซึ่งเท่ากับ 2935 พันตันข้าวสาร ได้แก่ ข้าวนึ่ง และต้นข้าวขาว

ตัวอย่างที่ 20

คะแนนสอบภาษาอังกฤษของนักเรียนชั้นมัธยมศึกษาปีที่ 6/1 ในโรงเรียนแห่งหนึ่ง จำนวนทั้งหมด 40 คน ซึ่งมีคะแนนเต็ม 100 คะแนน แสดงได้ดังนี้

96	78	80	76	84	77	74	85	65	69
82	53	45	67	58	54	56	62	56	54
43	48	49	50	60	65	54	51	55	60
65	66	75	98	97	63	92	94	76	78

จงหา

1) เปอร์เซ็นไทล์ที่ 25 เปอร์เซ็นไทล์ที่ 50 เปอร์เซ็นไทล์ที่ 75 และเปอร์เซ็นไทล์ที่ 80 ของคะแนนสอบวิชาภาษาอังกฤษของนักเรียนห้องนี้

2) คะแนนที่มีนักเรียนประมาณร้อยละ 25 ของห้องได้คะแนนมากกว่า

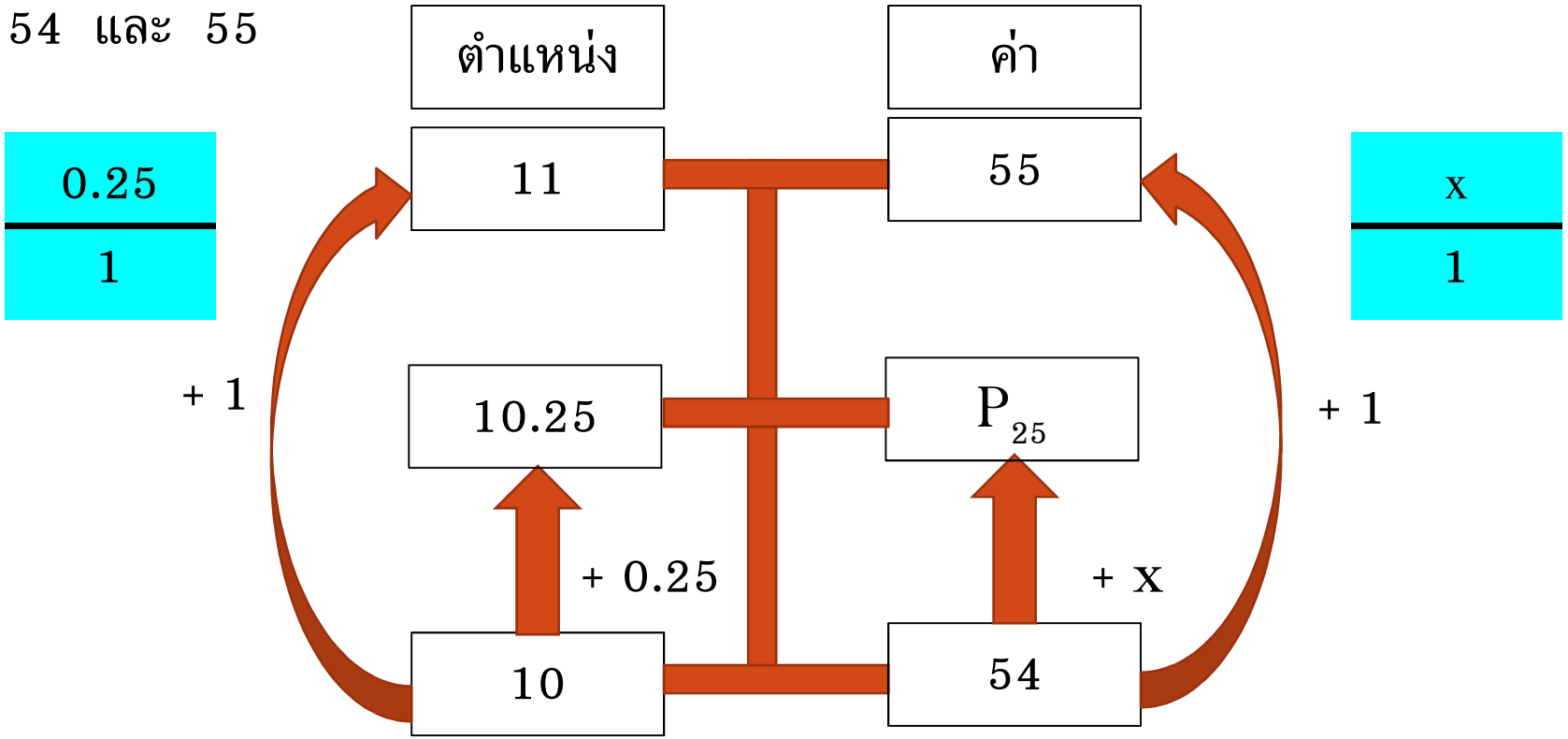
3) คะแนนที่มีนักเรียนประมาณหนึ่งในห้าของห้องได้คะแนนสูงกว่า

เรียงคะแนนสอบวิชาภาษาอังกฤษของนักเรียน 40 คน จากน้อยไปมาก ได้ดังนี้

43	45	48	49	50	51	53	54	54	54
55	56	56	58	60	60	62	63	65	65
65	66	67	69	74	75	76	76	77	78
78	80	82	84	85	92	94	96	97	98

1) เนื่องจาก P_{25} อยู่ในตำแหน่งที่ $\frac{25}{100}(40+1) = 10.25$

ดังนั้น P_{25} อยู่ระหว่างข้อมูลในตำแหน่งที่ 10 และตำแหน่งที่ 11 ซึ่งมีค่าอยู่ระหว่าง 54 และ 55



$$\frac{0.25}{1} = \frac{x}{1}$$

ดังนั้น $x = 0.25 \times 1 = 0.25$ นั่นคือ $P_{25} = 54 + 0.25 = 54.25$

เนื่องจาก P_{50} อยู่ในตำแหน่งที่ $\frac{50}{100}(40+1) = 20.5$

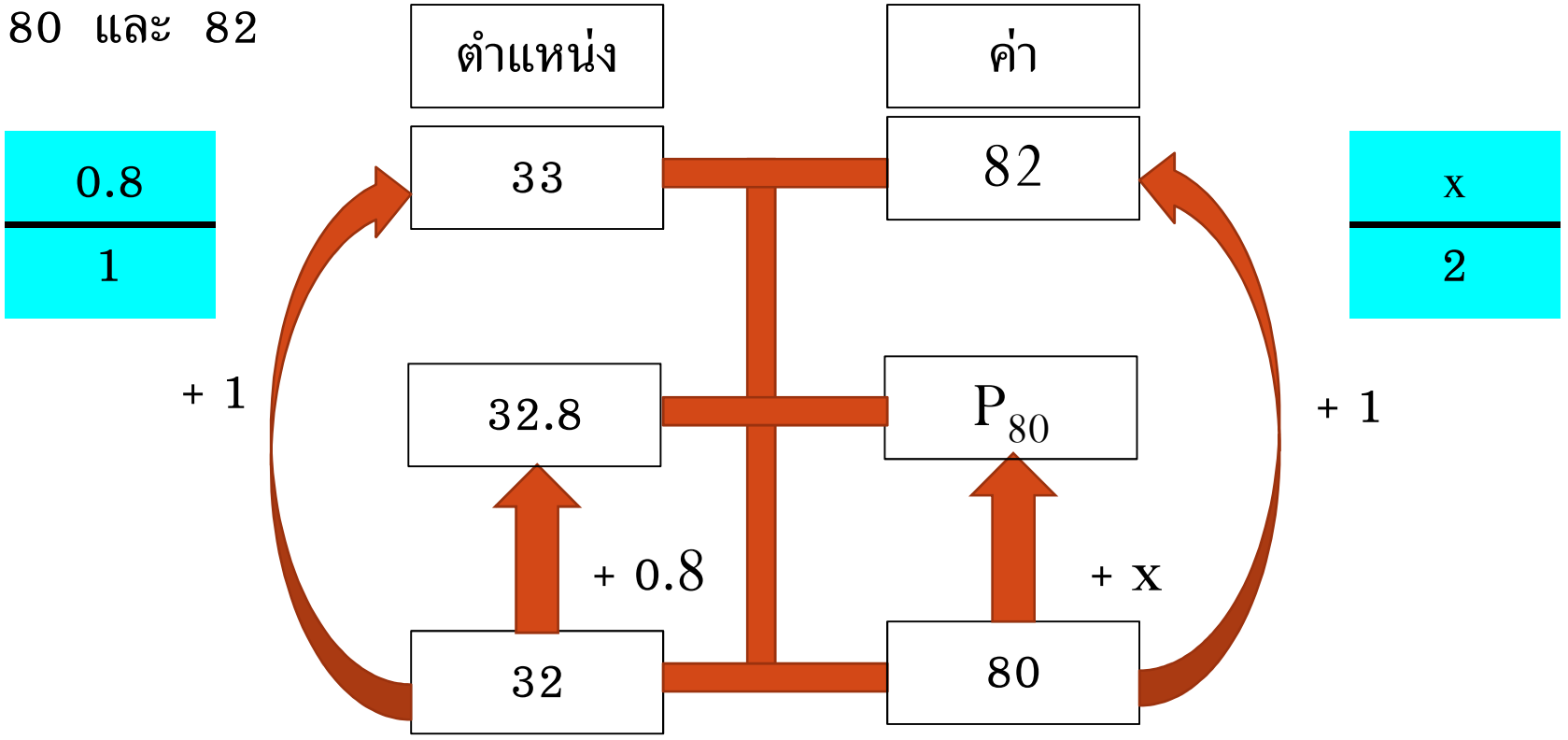
ดังนั้น P_{50} อยู่ระหว่างข้อมูลในตำแหน่งที่ 20 และตำแหน่งที่ 21 ซึ่งมีค่าอยู่ระหว่าง 65 และ 65 นั่นคือ $P_{50} = 65$

เนื่องจาก P_{75} อยู่ในตำแหน่งที่ $\frac{75}{100}(40+1) = 30.75$

ดังนั้น P_{75} อยู่ระหว่างข้อมูลในตำแหน่งที่ 30 และตำแหน่งที่ 31 ซึ่งมีค่าอยู่ระหว่าง 78 และ 78 นั่นคือ $P_{75} = 78$

1) เนื่องจาก P_{80} อยู่ในตำแหน่งที่ $\frac{80}{100}(40+1) = 32.8$

ดังนั้น P_{80} อยู่ระหว่างข้อมูลในตำแหน่งที่ 32 และตำแหน่งที่ 33 ซึ่งมีค่าอยู่ระหว่าง 80 และ 82



$$\frac{0.8}{1} = \frac{x}{2}$$

ดังนั้น $x = 0.8 \times 2 = 1.6$ นั่นคือ $P_{80} = 80 + 1.6 = 81.6$

จะได้ว่าเปอร์เซ็นไทล์ที่ 25 เปอร์เซ็นไทล์ที่ 50 เปอร์เซ็นไทล์ที่ 75 และเปอร์เซ็นไทล์ที่ 80 ของคะแนนสอบวิชาภาษาอังกฤษของนักเรียนห้องนี้ คือ 54.25, 65, 70 และ 81.6 คะแนน ตามลำดับ

2) คะแนนที่มีนักเรียนประมาณร้อยละ 25 ของห้องได้คะแนนต่ำกว่า คือ P_{25} ซึ่งเท่ากับ 54.25 คะแนน

3) คะแนนที่มีนักเรียนประมาณหนึ่งในห้าของห้องได้คะแนนสูงกว่า คือ P_{80} ซึ่งเท่ากับ 81.6 คะแนน

แบบฝึกหัดที่ 3.5

1. โรงไฟฟ้าพลังน้ำเขื่อนขนาดใหญ่จำนวน 15 แห่ง มีกำลังผลิต (เมกะวัตต์) ในเดือนมกราคม พ.ศ. 2562 ดังนี้

เขื่อน	กำลังผลิต (เมกะวัตต์)
ภูมิพล	779.20
สิริกิติ์	500.00
อุบลรัตน์	25.20
สิรินธร	36.00
จุฬาภรณ์	40.00
ศรีนครินทร์	720.00
วชิราลงกรณ์	300.00
ท่าทุ่งนา	39.00
แก่งกระจาน	19.00
บางลาง	84.00
รัชชประภา	240.00
ปากมูล	136.00
เจ้าพระยา	12.00
แควน้อยบำรุงแดน	30.00
แม่กลอง	12.00

จงหา

1) ควอร์ไทล์ที่ 1 ควอร์ไทล์ที่ 2 และควอร์ไทล์ที่ 3 ของข้อมูลชุดนี้

2) กำลังผลิตที่มีโรงไฟฟ้าพลังน้ำเขื่อนขนาดใหญ่จำนวนประชากรครั้งหนึ่งมีกำลังผลิตได้น้อยกว่า

3) โรงไฟฟ้าพลังน้ำเขื่อนขนาดใหญ่ที่มีกำลังผลิตมากกว่าควอร์ไทล์ที่ 3

2. โรงพยาบาลแห่งหนึ่งบันทึกจำนวนทารกแรกเกิดตั้งแต่เดือนมกราคม – ธันวาคม พ.ศ. 2562 ได้ข้อมูลดังนี้

เดือน	จำนวนทารกแรกเกิด (คน)
มกราคม	305
กุมภาพันธ์	289
มีนาคม	313
เมษายน	342
พฤษภาคม	311
มิถุนายน	324
กรกฎาคม	345
สิงหาคม	341
กันยายน	353
ตุลาคม	329
พฤศจิกายน	304
ธันวาคม	324

จงหา

1) ควอร์ไทล์ที่ 1 และควอร์ไทล์ที่ 3 ของข้อมูลชุดนี้

2) เดือนที่มีจำนวนทารกแรกเกิดน้อยกว่าควอรีไทล์ที่ 1

3) เดือนที่มีจำนวนทารกแรกเกิดมากกว่าควอร์ไทล์ที่ 3

3. จากการสำรวจของสำนักงานคณะกรรมการข้าราชการพลเรือน (ก.พ.) และสำนักงานสถิติแห่งชาติใน พ.ศ. 2556 พบว่า เงินเดือนของพนักงานใหม่แรกบรรจุ จำแนกตามวุฒิการศึกษาและตำแหน่ง แสดงได้ดังตารางต่อไปนี้

ตำแหน่ง	เงินเดือน (บาท)	
	ปริญญาตรี	ปริญญาโท/เอก
เจ้าพนักงานธุรการ เจ้าหน้าที่ธุรการ และพนักงานพิมพ์ดีด	12166	-
เจ้าพนักงานการเงินและบัญชี เจ้าหน้าที่บัญชีและเจ้าที่การเงิน	13184	19940
นักวิเคราะห์นโยบายและแผนและนักวางแผน	14308	22643
นักทรัพยากรบุคคล เจ้าหน้าที่ฝึกอบรวม เจ้าหน้าที่ทรัพยากรบุคคล และเจ้าหน้าที่วิเทศสัมพันธ์	13219	20957
นักประชาสัมพันธ์ และเจ้าหน้าที่ประชาสัมพันธ์	12760	18028
นักวิชาการคอมพิวเตอร์และนักเขียนโปรแกรม	15263	21342
นิติกร และนักกฎหมาย	16000	23823

ตำแหน่ง	เงินเดือน (บาท)	
	ปริญญาตรี	ปริญญาโท/เอก
เศรษฐกร และนักเศรษฐศาสตร์	15043	21050
วิศวกรเหมืองแร่ วิศวกรไฟฟ้า และและเครื่องกล	16986	24163
นักสำรวจ (ปิโตรเลียม) และวิศวกรปิโตรเลียม	17829	26686
สถาปนิก	18266	28677
นักทรัพยากรธรรมชาติ	18039	26555
นักวิชาการขนส่ง และนักโลจิสติกส์	15930	23125
นักวิทยาศาสตร์ และนักเคมี	16138	23454
แพทย์	-	63082
ทันตแพทย์	-	56807
พยาบาล	16487	39526
เภสัชกร	20003	33050
นักเทคนิคการแพทย์	16063	32505
นักรังสีการแพทย์	16267	37871
นักโภชนาการ	14973	30879

1) เงินเดือนของพนักงานใหม่แรกบรรจุที่มีวุฒิการศึกษาตรีหรือวุปริญญาโท/เอก มีการกระจายมากกว่ากัน

2) สำหรับพนักงานใหม่แรกบรรจุที่มีวุฒิปริญญาโท/เอก มีตำแหน่งใดบ้างที่ได้เงินเดือนน้อยกว่าเปอร์เซ็นไทล์ที่ 30

แบบฝึกหัดท้ายบท

1. ยอดชำระเงิน (บาท) ของลูกค้าที่มาซื้อของในซูเปอร์มาร์เก็ตของห้างสรรพสินค้าแห่งหนึ่ง จำนวน 50 คน แสดงได้ดังนี้

217	352	592	156	421	261	508	323	236	430
301	291	439	524	287	460	341	137	369	262
402	316	89	372	201	423	417	468	248	672
209	409	347	446	387	410	336	360	581	112
456	281	546	256	391	182	454	50	446	352

1) จงเขียนตารางความถี่พร้อมทั้งแสดงความถี่สะสม คาวมถี่สัมพัทธ์ และความถี่สะสมสัมพัทธ์ของข้อมูลชุดนี้ โดยกำหนดอันตรภาคชั้นเป็น ต่ำกว่า 100, 100 – 199, 200 – 299, 300 – 399, 400 – 499, 500 - 599 และ 600 - 699

2) ลูกค้ามียอดชำระเงินอยู่ในอันตรภาคใดมากที่สุด

3) ลูกค้าที่มียอดชำระเงินต่ำกว่า 100 บาท มีจำนวนมากหรือน้อยกว่าลูกค้าที่มียอดชำระเงนตั้งแต่ 600 - 699 บาท

4) ลูกค้าที่มียอดชำระเงินตั้งแต่ 400 บาทขึ้นไป คิดเป็นร้อยละเท่าใดของจำนวนลูกค้า 50 คนนี้

5) ลูกค้าที่มียอดชำระเงินตั้งแต่ 200 บาท แต่น้อยกว่า 500 บาท คิดเป็นร้อยละเท่าใดของจำนวนลูกค้า 50 คนนี้

2. คะแนนสอบวิชาคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ห้องหนึ่ง แสดงด้วยตารางความถี่ได้ดังนี้

คะแนน	จำนวนนักเรียน (คน)
30 – 39	2
40 – 49	0
50 – 59	6
60 – 69	6
70 – 79	10
80 – 89	13
90 – 99	8
รวม	45

ถ้ากำหนดเกณฑ์ในการแบ่งคะแนน ดังนี้

เกรด	คะแนน
4	90 – 100
3	80 – 89
2	70 – 79
1	50 – 69
0	ต่ำกว่า 50

1) จงเขียนตารางความถี่ของข้อมูลชุดนี้โดยกำหนดอันตรภาคชั้นเป็นระดับคะแนนที่กำหนดในเกณฑ์ข้างต้น

2) นักเรียนห้องนี้ได้เกรดใดมากที่สุด

3) นักเรียนที่ได้ตั้งแต่เกรด 3 ขึ้นไป คิดเป็นร้อยละเท่าใดของจำนวนทั้งหมด

3. ร้านค้าออนไลน์แห่งหนึ่งมีนโยบายการรับประกันความพึงพอใจโดยลูกค้าสามารถส่งคืนสินค้าที่ยังไม่ได้ใช้งานภายใน 7 วัน นับจากวันที่ได้รับคืนสินค้า เพื่อรับเงินคืนเต็มจำนวน ถ้าข้อมูลจำนวนสินค้า (ชิ้น) ที่ลูกค้าส่งคืนในแต่ละวันในระยะเวลา 30 วัน แสดงได้ดังนี้

4	9	5	6	5	8	22	3	8	4
8	2	7	19	12	3	5	16	3	6
5	15	9	14	4	13	7	2	9	7

1) จงเขียนตารางความถี่ของข้อมูลชุดนี้พร้อมทั้งแสดงขอบล่างและขอบบนของแต่ละชั้น โดยกำหนดให้จำนวนอันตรภาคชั้นเท่ากับ 5 ชั้น ค่าเริ่มต้นเท่ากับ 0 ชิ้น และค่าสุดท้ายเท่ากับ 25 ชิ้น

2) จงเขียนฮิสโทแกรมของข้อมูลชุดนี้

3) จงเขียนแผนภาพจุดของข้อมูลชุดนี้

4) จงเขียนแผนภาพกล่องของข้อมูลชุดนี้

4. สายการบินแห่งหนึ่งได้สำรวจความพึงพอใจของผู้โดยสารต่อการให้บริการของพนักงานต้อนรับบนเครื่องบิน ซึ่งมีคะแนนเต็ม 10 คะแนน โดยสำรวจจากผู้โดยสารที่สุ่มมาจำนวน 20 คน ได้ผลสำรวจโดยเรียงคะแนนจากน้อยไปมากดังนี้

1 3 5 5 7 8 8 8 8 8 8 9 9 9 9 9 10 10 10 10

1) จงหาควอร์ไทล์ที่ 1 ควอร์ไทล์ที่ 2 และควอร์ไทล์ที่ 3 ของข้อมูลชุดนี้

2) ข้อมูลชุดนี้ค่านอกเกณฑ์หรือไม่ ถ้ามีคือค่าใด

3) จงเขียนแผนภาพกล่องเพื่อนำเสนอข้อมูลชุดนี้ พร้อมทั้งอธิบายลักษณะการกระจายของข้อมูลชุดนี้

5. กำหนดข้อมูลของประชากรในแต่ละชุดดังต่อไปนี้

ชุด ก : 5 5 5 5 5 6 6 6 7 7 8 8 9 9 10 10

ชุด ข : 1 1 1 1 2 2 2 3 3 4 30

จงหาค่าเฉลี่ยเลขคณิต มัธยฐาน และฐานนิยมของข้อมูลแต่ละชุด

6. กำหนดข้อมูลของตัวอย่างในแต่ละชุดดังต่อไปนี้

ชุด ก : 2 3 3 5 5 5 7 7 8 10 14 14 19

ชุด ข : 0.9 1.2 1.7 2.1 2.5 2.8 3.2 3.3 3.7 4.8 5.7

ชุด ค : 59 73 82 87 87 90

จงหา

1) ค่าเฉลี่ยเลขคณิต มัธยฐาน และฐานนิยมของข้อมูลแต่ละชุด

2) พิสัย พิสัยระหว่างควอร์ไทล์ ส่วนเบี่ยงเบนมาตรฐาน และความแปรปรวนของข้อมูลแต่ละชุด

3) ข้อมูลชุดใดมีการกระจายมากที่สุด

7. ศิริวิทย์ได้รับผลการสอบวิชาภาษาไทย คณิตศาสตร์ ภาษาอังกฤษ และคอมพิวเตอร์ พบว่าค่าเฉลี่ยเลขคณิตของคะแนนสอบวิชาคณิตศาสตร์และวิชาภาษาไทยเท่ากับ 21.5 คะแนน และค่าเฉลี่ยเลขคณิตของคะแนนสอบวิชาภาษาอังกฤษและวิชาคอมพิวเตอร์เท่ากับ 28.5 คะแนน จงหาค่าเฉลี่ยเลขคณิตของคะแนนสอบทั้งสี่วิชานี้

8. ถ้าค่าเฉลี่ยเลขคณิตของอายุนักเรียนชั้นมัธยมศึกษาปีที่ 3, 4 และ 5 ของโรงเรียนแห่งหนึ่ง เท่ากับ 15, 17 และ 18 ปี ตามลำดับ และจำนวนนักเรียนชั้นมัธยมศึกษาปีที่ 3, 4 และ 5 ของโรงเรียนแห่งนี้เท่ากับ 60, 50 และ 40 คน ตามลำดับ จงหาค่าเฉลี่ยเลขคณิตของอายุนักเรียนชั้นมัธยมศึกษาปีที่ 3 - 5 ของโรงเรียนแห่งนี้

9. ข้อมูลชุดหนึ่งประกอบด้วยจำนวนเต็ม 5 ตำนวน โดยค่าเฉลี่ยเลขคณิตของข้อมูลชุดนี้เท่ากับ 360 เมื่อเรียงข้อมูลจากมากไปน้อย ปรากฏว่าข้อมูล 2 ตัวสุดท้าย คือ 102 และ 99 ตามลำดับ จงหาค่าที่มากที่สุดที่เป็นไปได้ของข้อมูลชุดนี้

10. ข้อมูลชุดหนึ่งประกอบด้วยจำนวนเต็ม 5 จำนวน โดยค่าเฉลี่ยเลขคณิต มัธยฐาน ฐานนิยม และพิสัยของข้อมูลชุดนี้เท่ากับ 5 จงหาข้อมูลของข้อมูลที่เป็นไปได้ทั้งหมด

11. ร้านค้าแห่งหนึ่งต้องการวิเคาะห์ข้อมูลการขายสินค้าของร้าน จึงได้บันทึกจำนวนสินค้า (ชิ้น) ที่ขายได้ในหนึ่งวัน โดยสุ่มเก็บข้อมูลเพียง 11 วัน ในหนึ่งเดือนที่ผ่านมา ได้ข้อมูลดังนี้

85 125 30 75 80 65 90 75 78 92 67

1) จงหาควอร์ไทล์ที่ 1 ควอร์ไทล์ที่ 2 และควอร์ไทล์ที่ 3 ของข้อมูลชุดนี้

2) จงหาพิสัยและพิสัยระหว่างควอร์ไทล์ของข้อมูลชุดนี้ และพิจารณาว่าค่าวัดการกระจายสัมบูรณ์ใดเหมาะสมำหรับใช้อธิบายการกระจายของข้อมูลชุดนี้ พร้อมทั้งให้เหตุผลประกอบ

3) จงหาเปอร์เซ็นไทล์ที่ 25 เปอร์เซ็นไทล์ที่ 50 เปอร์เซ็นไทล์ที่ 75 และเปอร์เซ็นไทล์ที่ 90 ของข้อมูลชุดนี้

12. เวลา (นาที) ที่ใช้ในการอ่านหนังสือในหนึ่งวันของนักเรียน 2 ห้องเรียน ที่สุ่มมาจำนวนห้องละ 10 คน แสดงได้ดังนี้

ห้อง 1	0	20	30	42	35	82	54	28	0	63
ห้อง 2	45	40	62	10	24	15	30	60	95	120

จงหาสัมประสิทธิ์การแปรผันของเวลาที่ใช้ในการอ่านหนังสือในหนึ่งวันของนักเรียนแต่ละห้องที่สุ่มมา พร้อมทั้งเปรียบเทียบการกระจายของเวลาที่ใช้ในการอ่านหนังสือในหนึ่งวันของนักเรียนทั้งสองห้อง

13. มานี ชูใจ และปิติ ได้ค่าขนมจากผู้ปกครอง โดยค่าเฉลี่ยเลขคณิตและส่วนเบี่ยงเบนมาตรฐานของค่าขนมของทั้งสามคนเท่ากับ 50 และ 0 บาท ตามลำดับ ถ้านำค่าขนมของมานะมาคำนวณด้วยจะได้ค่าเฉลี่ยเลขคณิตใหม่เป็น 45 บาท จงหาส่วนเบี่ยงเบนมาตรฐานของค่าขนมของทั้งสี่คน

14. กำหนดให้ข้อมูลของประชากรจำนวน 2 ชุด มีลักษณะดังต่อไปนี้

ชุด ก: ประกอบด้วยจำนวนเต็มที่แตกต่างกัน 5 จำนวน โดยทั้ง 5 จำนวนนี้อยู่ระหว่าง 0 และ 10

ชุด ข: ประกอบด้วยจำนวนเต็มที่แตกต่างกัน 5 จำนวน โดยทั้ง 5 จำนวนนี้อยู่ระหว่าง 10 และ 20

1) จงยกตัวอย่างข้อมูลชุด ก และ ข ที่ส่วนเบี่ยงเบนมาตรฐานของข้อมูลชุด ก มากกว่าส่วนเบี่ยงเบนมาตรฐานของข้อมูลชุด ข พร้อมทั้งคำนวณประกอบ

2) มีโอกาสที่ส่วนเบี่ยงเบนมาตรฐานของข้อมูลแต่ละชุดจะเป็น 0 หรือไม่ เพราะเหตุใด

3) จงยกตัวอย่างข้อมูลชุด ก และ ข ที่ส่วนเบี่ยงเบนมาตรฐานของข้อมูลชุด ก เท่ากับส่วนเบี่ยงเบนมาตรฐานของข้อมูลชุด ข

15. กำหนดข้อมูลของปราะชากรชุดหนึ่งเรียงลำดับจากน้อยไปมากได้ดังนี้

1 2 x 3 3 y 6

ถ้าค่าเฉลี่ยเลขคณิตและส่วนเบี่ยงเบนมาตรฐานของข้อมูลชุดนี้เท่ากับ 3 และ $\frac{4\sqrt{7}}{7}$ ตามลำดับ จงหาค่า x และ y

16. ข้อมูลของตัวอย่างชุดหนึ่งมี 20 ตัว โดยค่าเฉลี่ยเลขคณิตและส่วนเบี่ยงเบนมาตรฐานของข้อมูลชุดนี้เท่ากับ 10 และ 2 ตามลำดับ ถ้าภายหลังพบว่ามีการบันทึกข้อมูลตัวหนึ่งผิดพลาดไปจาก 12 บันทึกผิดเป็น 8 จงหาค่าเฉลี่ยเลขคณิตและความแปรปรวนที่ถูกต้องของข้อมูลชุดนี้

17. ข้อสอบวิชาหนึ่งมีคะแนนเต็ม 80 คะแนน โดยคะแนนของผู้เข้าสอบ 10 คน แสดงได้ดังนี้

25 62 59 38 38 41 69 72 26 44

1) ถ้าเกณฑ์ในการสอบผ่านคือต้องได้คะแนนไม่ต่ำกว่าร้อยละ 70 ของคะแนนเต็ม จงหาคะแนนต่ำสุดของผู้ที่สอบผ่าน

2) ถ้าเกณฑ์ในการสอบผ่านคือต้องได้คะแนนไม่ต่ำกว่าเปอร์เซ็นไทล์ที่ 70 จงหาคะแนนต่ำสุดของผู้ที่สอบผ่าน

18. ผลการทดสอบความสามารถในการคิดอย่างมีวิจารณญาณของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ของโรงเรียนแห่งหนึ่ง จำนวนทั้งหมด 40 คน ซึ่งมีคะแนนเต็ม 100 คะแนน โดยเรียงคะแนนจากน้อยไปมาก แสดงได้ดังนี้

43	45	48	49	50	51	53	54	54	54
55	56	56	58	60	60	62	63	65	65
65	66	67	69	74	75	76	76	77	78
78	80	82	84	85	92	94	96	97	98

1) จงหาคะแนนที่มีนักเรียนประมาณครึ่งหนึ่งของชั้นได้คะแนนต่ำกว่า

2) จงหาคะแนนที่มีนักเรียนประมาณหนึ่งในสี่ของชั้นได้คะแนนสูงกว่า

3) จงหาคะแนนที่มีนักเรียนประมาณหกในสิบของชั้นได้คะแนนต่ำกว่า

4) ถ้าโรงเรียนแห่งนี้จัดกิจกรรมพัฒนาทักษะการคิดอย่างมีวิจารณญาณให้กับนักเรียนที่ได้คะแนนน้อยกว่าเปอร์เซ็นไทล์ที่ 25 จงหาจำนวนนักเรียนที่ต้องเข้าร่วมกิจกรรมนี้