สถิติ

16 Apr 2020

สารบัญ

สถิติภาคบรรยาย

สถิติ จะแปลว่า “ข้อมูล” ก็ได้ หรือจะแปลว่า “วิธีจัดการข้อมูล” ตามระเบียบวิธีการทางสถิติ ก็ได้
สถิติมีหลายชนิด ตามลักษณะการนำไปใช้ เช่น

- สถิติเชิงพรรณนา = สถิติที่มุ่งอธิบายภาพรวมของข้อมูล
- สถิติเชิงอนุมาน = สถิติที่ศึกษากลุ่มตัวอย่างเล็กๆ เพื่อเดาข้อสรุปของข้อมูลทั้งหมด
- สถิติเพื่อการตัดสินใจ = สถิติชนิดนี้ จะไม่ต้องการภาพรวมของข้อมูล แต่จะมุ่งเน้นเพื่อตัดสินใจบางเรื่อง โดยจะกำหนดก่อนว่าต้องใช้ค่าสถิติใดในการตัดสินใจ แล้วจึงเก็บข้อมูลเท่าที่จำเป็น

<u>ระเบียบวิธีการทางสถิติ</u> ประกอบด้วย การเก็บข้อมูล → นำเสนอข้อมูล → วิเคราะห์ → ตีความ

<u>การเก็บข้อมูล</u> สามารถเก็บจากทะเบียนประวัติ จากการสำรวจ จากการทดลอง หรือ จากการสังเกต
โดยเราแบ่งประเภทข้อมูล ตามวิธีเก็บได้เป็น 2 ชนิด

- ข้อมูลปฐมภูมิ = ข้อมูลที่เก็บจากต้นกำเนิดข้อมูลโดยตรง
- ข้อมูลทุติยภูมิ = ข้อมูลที่เก็บจากผลการเก็บข้อมูลของคนอื่นอีกที เช่น เก็บจากเอกสาร หรือทะเบียนที่มีอยู่

ซึ่งก่อนเก็บข้อมูล เราต้องเลือกอีกว่าจะเก็บจาก “ประชากร” หรือเก็บจาก “ตัวอย่าง”

- ประชากร = ข้อมูล “ทั้งหมด”
- ตัวอย่าง = ข้อมูล “บางส่วน”

การเก็บข้อมูลจากประชากร จะได้ผลครบถ้วนสมบูรณ์กว่าเก็บจากตัวอย่าง แต่ก็ใช้เวลามากกว่า
ในชีวิตจริง เรามักนิยมเก็บข้อมูลจากจากตัวอย่าง
โดยจะมีวิธีเลือกตัวอย่างตามหลักสถิติ เพื่อให้ตัวอย่างที่เลือก สามารถเป็นตัวแทนของประชากรได้ใกล้เคียงที่สุด

<u>การนำเสนอข้อมูล</u> คือ การเอาข้อมูลมาจัดเป็นรูปแบบที่เข้าใจง่าย ซึ่งแบ่งได้เป็น 2 แบบ คือ

- ไม่เป็นแบบแผน ได้แก่ การนำเสนอในรูปบทความ หรือ แบบกึ่งตาราง
- เป็นแบบแผน ได้แก่ การนำเสนอด้วยตาราง แผนภูมิ หรือ กราฟ

<u>การวิเคราะห์ข้อมูล</u> คือ การหาข้อสรุปเกี่ยวกับลักษณะต่างๆ (ค่าเฉลี่ย, มัธยฐาน, ฐานนิยม, ฯลฯ) ของข้อมูล
ค่าที่วิเคราะห์ได้ จะมีชื่อเรียก 2 ชื่อ ขึ้นกับว่าเอาข้อมูลระดับไหนมาวิเคราะห์

- พารามิเตอร์ = ค่าที่ได้จาก “ประชากร”
- ค่าสถิติ = ค่าที่ได้จาก “ตัวอย่าง”

ปกติ เราจะไม่วิเคราะห์ประชากร แต่เราจะวิเคราะห์ตัวอย่าง เพื่อหาค่าสถิติแล้วนำมาประมาณพารามิเตอร์
โดยข้อมูลที่นำมาวิเคราะห์ จะแบ่งได้เป็น 2 ประเภท

- ข้อมูลเชิงปริมาณ คือ ข้อมูลที่บอกปริมาณ เช่น อายุ ส่วนสูง น้ำหนัก
- ข้อมูลเชิงคุณภาพ คือ ข้อมูลที่แสดงคุณสมบัติ เช่น เพศ กรุ๊ปเลือด ข้อมูลประเภทนี้ จะไม่สามารถ บวก ลบ คูณ หาร หรือเทียบมากกว่าน้อยกว่า ได้ จึงหาไม่สามารถหาค่าเฉลี่ย หรือ มัธยฐาน ได้

<u>การตีความข้อมูล</u> คือ การหาข้อสรุปจากค่าที่ได้จากการวิเคราะห์

<u>แบบฝึกหัด</u>

1. ข้อใดถูกต้อง
    1. เบอร์โทรศัพท์ เป็นข้อมูลเชิงปริมาณ
    2. การเก็บข้อมูลของชุมชน จากที่ว่าการเขต จัดเป็นการเก็บข้อมูลแบบปฐมภูมิ
    3. การหาค่าเฉลี่ยของข้อมูล เป็นส่วนหนึ่งของการนำเสนอข้อมูล
    4. สถิติเชิงอนุมาน คือ สถิติที่มุ่งอธิบายลักษณะกว้างๆของข้อมูล
    5. พารามิเตอร์ คือ ค่าที่ได้จากการนำกลุ่มตัวอย่างมาคำนวณ
    6. ประโยชน์ของการนำเสนอข้อมูล คือ ทำให้เข้าใจข้อมูลได้ง่ายขึ้น

2. ข้อต่อไปนี้มีผลกระทบต่อความถูกต้องของการตัดสินใจโดยใช้สถิติ <u>ยกเว้น</u>ข้อใด [O-NET 52/34]
    1. ข้อมูล 2. สารสนเทศ 3. ข่าวสาร 4. ความเชื่อ

3. ข้อใดเป็นขั้นตอนหนึ่งของการสำรวจความคิดเห็น [O-NET 57/28]
    1. ตั้งสมมุติฐานของปัญหาที่ทำการสำรวจ
    2. กำหนดขอบเขตของการสำรวจ
    3. ประมาณการค่าใช้จ่ายในการสำรวจความคิดเห็น
    4. คัดเลือกผู้เก็บข้อมูลการสำรวจ
    5. นำผลการสำรวจความคิดเห็นไปใช้ประโยชน์

4. ข้อใด<u>ไม่</u>อยู่ในขั้นตอนของการสำรวจความคิดเห็น [O-NET 56/30]
    1. กำหนดขอบเขตของการสำรวจ 2. กำหนดวิธีเลือกตัวอย่าง
    3. สร้างแบบสำรวจความคิดเห็น 4. ประมวลผลและวิเคราะห์ผลการสำรวจ
    5. เผยแพร่ผลการสำรวจความคิดเห็น

5. ในการใช้สถิติเพื่อการตัดสินใจและวางแผน สำหรับเรื่องที่จำเป็นต้องมีการใช้ข้อมูลและสารสนเทศ ถ้าขาดข้อมูลและสารสนเทศดังกล่าว ผู้ตัดสินใจควรทำขั้นตอนใดก่อน [O-NET 53/36]
    1. เก็บรวบรวมข้อมูล 2. เลือกวิธีวิเคราะห์ข้อมูล
    3. เลือกวิธีเก็บรวบรวมข้อมูล 4. กำหนดข้อมูลที่จำเป็นต้องใช้

6. ข้อใดต่อไปนี้<u>เป็นเท็จ</u> [O-NET 52/25]
    1. สถิติเชิงพรรณนาคือสถิติของการวิเคราะห์ข้อมูลขั้นต้นที่มุ่งอธิบายลักษณะกว้างๆของข้อมูล
    2. ข้อมูลที่เป็นหมายเลขที่ใช้เรียกสายรถโดยสารประจำทางเป็นข้อมูลเชิงคุณภาพ
    3. ข้อมูลปฐมภูมิคือข้อมูลที่ผู้ใช้เก็บรวบรวมจากแหล่งข้อมูลโดยตรง
    4. ข้อมูลที่นักเรียนรวบรวมจากรายงานต่างๆที่ได้จากหน่วยงานราชการเป็นข้อมูลปฐมภูมิ

7. ครูสอนวิทยาศาสตร์มอบหมายให้นักเรียน 40 คน ทำโครงงานตามความสนใจ หลังจากตรวจรายงานโครงงานของทุกคนแล้ว ผลสรุปเป็นดังนี้

ผลการประเมิน	จำนวนโครงงาน
ดีเยี่ยม	3
ดี	20
พอใช้	12
ต้องแก้ไข	5

ข้อมูลที่เก็บรวบรวม เพื่อให้ได้ผลสรุปข้างต้นเป็นข้อมูลชนิดใด [O-NET 53/27]

1. ข้อมูลปฐมภูมิ เชิงปริมาณ
2. ข้อมูลทุติยภูมิ เชิงปริมาณ
3. ข้อมูลปฐมภูมิ เชิงคุณภาพ
4. ข้อมูลทุติยภูมิ เชิงคุณภาพ

การแจกแจงความถี่

"ตารางแจกแจงความถี่" คือ ตารางที่บอกว่า มีข้อมูลกี่ตัวตกอยู่ในแต่ละช่วง

ความถี่ (f)
เป็นค่าที่บอกว่ามีข้อมูลกี่ตัว ตกอยู่
ในแต่ละช่วง ในที่นี้คือจำนวน
นักเรียนในแต่ละช่วงคะแนน

ค่าของข้อมูล (x)
ในที่นี้คือคะแนนสอบ
ของนักเรียนแต่ละคน

คะแนนสอบ	จำนวนนักเรียน (คน)
1 - 10	3
11 - 20	12
21 - 30	15
31 - 40	24
41 - 50	6

→ อันตรภาคชั้น
→ เรียกสั้นๆว่า "ชั้น" หมายถึง
→ แต่ละแถวของช่วงข้อมูล
→ เช่น ในตารางนี้ จะมี 5 ชั้น

บางตารางจะใจดี บวกจำนวนข้อมูลมาให้ข้างล่าง บางตารางก็มีช่อง "ความถี่สะสม" มาให้ ค่าพวกนี้เราคำนวณเองได้

คะแนนสอบ	จำนวนนักเรียน (คน)	ความถี่สะสม (F)
1 - 10	3	3
11 - 20	12	15
21 - 30	15	30
31 - 40	24	54
41 - 50	6	60
	60	

ความถี่สะสม (F)
ได้จากการนำช่อง "ความถี่" มาบวก
สะสม โดยเริ่มสะสมตั้งแต่ชั้นแรก

จำนวนข้อมูลทั้งหมด (N)

นอกจากนี้ ยังมีช่องอื่นๆที่ไม่ค่อยจะได้ใช้อีก ดังนี้

- ความถี่<u>สัมพัทธ์</u> $= \dfrac{\text{ความถี่}}{\text{จำนวนข้อมูลทั้งหมด}} = \dfrac{f}{N}$
- ความถี่สะสม<u>สัมพัทธ์</u> $= \dfrac{\text{ความถี่สะสม}}{\text{จำนวนข้อมูลทั้งหมด}} = \dfrac{F}{N}$

สัมพัทธ์ $\rightarrow \div N$

- <u>ร้อยละ</u>ของความถี่สัมพัทธ์ $=$ ความถี่สัมพัทธ์ $\times 100$
- <u>ร้อยละ</u>ของความถี่สะสมสัมพัทธ์ $=$ ความถี่สะสมสัมพัทธ์ $\times 100$

ร้อยละ $\rightarrow \times 100$

$\div N$ $\times 100$

คะแนนสอบ	จำนวนนักเรียน (คน)	ความถี่สะสม (F)	ความถี่สัมพัทธ์	ความถี่สะสมสัมพัทธ์	ร้อยละของความถี่สัมพัทธ์	ร้อยละของความถี่สะสมสัมพัทธ์
1 - 10	3	3	$3/60 = 0.05$	$3/60 = 0.05$	$0.05 \times 100 = 5$	$0.05 \times 100 = 5$
11- 20	12	15	$12/60 = 0.20$	$15/60 = 0.25$	$0.20 \times 100 = 20$	$0.25 \times 100 = 25$
21 - 30	15	30	$15/60 = 0.25$	$30/60 = 0.50$	$0.25 \times 100 = 25$	$0.50 \times 100 = 50$
31 - 40	24	54	$24/60 = 0.40$	$54/60 = 0.90$	$0.40 \times 100 = 40$	$0.90 \times 100 = 90$
41 - 50	6	60	$6/60 = 0.10$	$60/60 = 1.00$	$0.10 \times 100 = 10$	$1.00 \times 100 = 100$
	60		1.00		100	

ความถี่สัมพัทธ์
รวมกันได้ 1 เสมอ

ร้อยละของความถี่สัมพัทธ์
รวมกันได้ 100 เสมอ

ในตารางแจกแจงความถี่ มีคำศัพท์ที่ต้องรู้ดังนี้

- ขอบล่าง $= \dfrac{\text{ค่าต่ำสุดของชั้น}+\text{ค่าสูงสุดของชั้นก่อนหน้า}}{2}$
- ขอบบน $= \dfrac{\text{ค่าสูงสุดของชั้น}+\text{ค่าต่ำสุดของชั้นถัดไป}}{2}$
- จุดกึ่งกลางชั้น $= \dfrac{\text{ค่าสูงสุดของอันตรภาคชั้น}+\text{ค่าต่ำสุดของอันตรภาคชั้น}}{2}$

  $= \dfrac{\text{ขอบบน}+\text{ขอบล่าง}}{2}$
- ความกว้างชั้น (I) $=$ ขอบบน $-$ ขอบล่าง

  $=$ ค่าข้อมูลตำแหน่งเดียวกัน ระหว่าง 2 ชั้นที่ติดกัน ลบกัน

คะแนนสอบ	ขอบล่าง	ขอบบน	จุดกึ่งกลางชั้น	ความกว้างชั้น (I)
1 - 10	0.5	$(10+11)/2 = 10.5$	$(1+10)/2 = 5.5$	$10.5 - 0.5 = 10$
11 - 20	$(10+11)/2 = 10.5$	$(20+21)/2 = 20.5$	$(11+20)/2 = 15.5$	$20.5 - 10.5 = 10$
21 - 30	$(20+11)/2 = 20.5$	$(30+31)/2 = 30.5$	$(21+30)/2 = 25.5$	$30.5 - 20.5 = 10$
31 - 40	$(30+31)/2 = 30.5$	$(40+41)/2 = 40.5$	$(31+40)/2 = 35.5$	$40.5 - 30.5 = 10$
41 - 50	$(40+41)/2 = 40.5$	50.5	$(41+50)/2 = 45.5$	$50.5 - 40.5 = 10$

หมายเหตุ: ขอบล่างของชั้นต่ำสุด จะคำนวณจากแนวโน้มความห่างของขอบบน

ขอบบนของชั้นสูงสุด จะคำนวณจากแนวโน้มความห่างของขอบล่าง

<u>แบบฝึกหัด</u>

1. จงเติมตารางแจกแจงความถี่ให้สมบูรณ์

   1. 27 1 14 8 30 20 8 27 12 14
      12 27 3 5 5 19 2 17 15 14

ค่าข้อมูล	ความถี่	ความถี่สะสม	ความถี่สัมพัทธ์	ความถี่สะสมสัมพัทธ์	ร้อยละของความถี่สัมพัทธ์	ร้อยละของความถี่สะสมสัมพัทธ์
1 - 10						
11 - 20						
21 - 30						

   2. 5 12 10 22 11 20 16 8 14 29
      20 6 16 12 9 18 18 15 34 34

ค่าข้อมูล	ความถี่	ความถี่สะสม	จุดกึ่งกลางชั้น	ความกว้างชั้น	ขอบล่าง	ขอบบน
1 - 10						
11 - 20						
21 - 30						
31 - 40						

3.

ค่าข้อมูล	ความถี่	ความถี่สะสม	จุดกึ่งกลางชั้น	ความกว้างชั้น	ขอบล่าง	ขอบบน
		2	5.5			
		5	11.5			
		16	17.5			
		20	23.5			

4.

ค่าข้อมูล	ความถี่	ความถี่สะสม	จุดกึ่งกลางชั้น	ความกว้างชั้น	ขอบล่าง	ขอบบน
		5	2.5			
		12	6.5			
	13			4		
		30		4		

2. ในการสำรวจอายุของคนในหมู่บ้านแห่งหนึ่งเป็นดังนี้

อายุ (ปี)	ความถี่ (คน)	ความถี่สัมพัทธ์
0 – 10	10	
11 – 20	25	
21 – 30	35	
31 – 40		x
41 – 50	40	
51 – 60	20	0.10
61 – 70	15	
71 – 80	3	
81 – 90	2	

ค่า x ในตารางแจกแจงความถี่สัมพัทธ์เท่ากับเท่าใด [O-NET 54/39]

3. ข้อมูลชุดหนึ่ง มีบางส่วนถูกนำเสนอในตารางต่อไปนี้

อันตรภาคชั้น	ความถี่	ความถี่สะสม	ความถี่สัมพัทธ์
2 - 6			
7 - 11		11	0.2
12 - 16		14	
17 - 21	6		0.3

ช่วงคะแนนใดเป็นช่วงคะแนนที่มีความถี่สูงสุด [O-NET 53/34]

4. ตารางแสดงน้ำหนักของนักเรียนจำนวน 50 คน เป็นดังนี้

น้ำหนัก (กิโลกรัม)	จำนวน (คน)
30 - 39	4
40 - 49	5
50 - 59	13
60 - 69	17
70 - 79	6
80 - 89	5

ข้อสรุปในข้อใดต่อไปนี้ ไม่ถูกต้อง [O-NET 49/1-27]

1. นักเรียนกลุ่มนี้ส่วนใหญ่มีน้ำหนัก 60 - 69 กิโลกรัม
2. นักเรียนที่มีน้ำหนักต่ำกว่า 50 กิโลกรัม มี 9 คน
3. นักเรียนที่มีน้ำหนักในช่วง 50 - 59 กิโลกรัม มี 26 %
4. นักเรียนที่มีน้ำหนักมากกว่า 80 กิโลกรัม มี 10 %

5. กำหนดให้ตารางแจกแจงความถี่สะสมของคะแนนของนักเรียนห้องหนึ่ง เป็นดังนี้

ช่วงคะแนน	ความถี่สะสม
30 - 39	1
40 - 49	11
50 - 59	18
60 - 69	20

ข้อสรุปในข้อใดต่อไปนี้ ถูกต้อง [O-NET 50/37]

1. นักเรียนที่ได้คะแนน 40 - 49 คะแนน มีจำนวน 27 %
2. นักเรียนส่วนใหญ่ได้คะแนน 60 - 69 คะแนน
3. นักเรียนที่ได้คะแนนมากกว่า 53 คะแนน มีจำนวนน้อยกว่า
นักเรียนที่ได้คะแนน 40 - 49 คะแนน
4. นักเรียนที่ได้คะแนนน้อยกว่า 47 คะแนน มีจำนวนมากกว่า
นักเรียนที่ได้คะแนนมากกว่า 50 คะแนน

6. จำนวนผู้ว่างงานทั่วประเทศในเดือนกันยายน ปี พ.ศ. 2551 มีจำนวนทั้งสิ้น 4.29 แสนคน ตารางเปรียบเทียบอัตราการว่างงานในเดือนกันยายน ปีพ.ศ. 2550 กับปีพ.ศ. 2551 เป็นดังนี้

พื้นที่สำรวจ	อัตราการว่างงานในเดือนกันยายน (จำนวนผู้ว่างงานต่อจำนวนผู้อยู่ในกำลังแรงงานคูณ 100)	
	ปีพ.ศ. 2550	ปีพ.ศ. 2551
ภาคใต้	1.0	1.0
ภาคตะวันออกเฉียงเหนือ	0.9	1.3
ภาคเหนือ	1.5	1.2
ภาคกลาง (ยกเว้นกรุงเทพมหานคร)	1.3	0.9
กรุงเทพมหานคร	1.2	1.2
ทั่วประเทศ	1.2	1.1

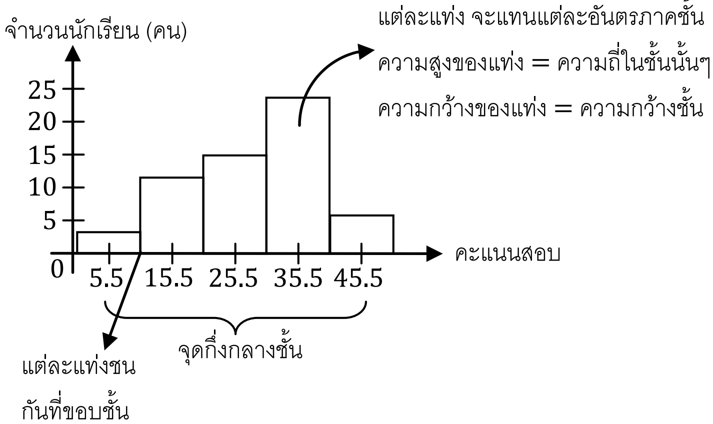
ข้อใดถูกต้องบ้าง [O-NET 53/35]

1. จำนวนผู้ว่างงานในภาคใต้ในเดือนกันยายนของปีพ.ศ. 2550 และของปีพ.ศ. 2551 เท่ากัน
2. จำนวนผู้อยู่ในกำลังแรงงานทั่วประเทศในเดือนกันยายน ปีพ.ศ. 2551 มีประมาณ 39 ล้านคน

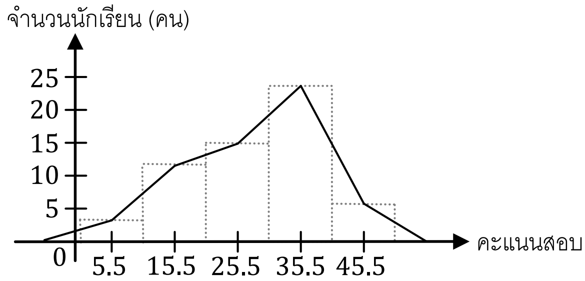
กราฟแจกแจงความถี่

ในเรื่องนี้จะพูดถึงวิธีการนำตารางแจกแจงความถี่ มาเขียนเป็นแผนภูมิให้อ่านง่าย
ก่อนอื่น เราสามารถนำตารางแจกแจงความถี่ มาเขียนเป็นแผนภูมิแท่งที่เรียกว่า “ฮิสโทแกรม” ได้ ดังนี้

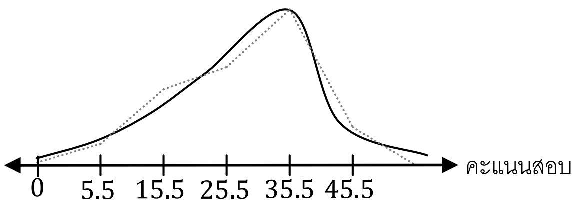
คะแนนสอบ	จำนวนนักเรียน (คน)
1 - 10	3
11 - 20	12
21 - 30	15
31 - 40	24
41 - 50	6



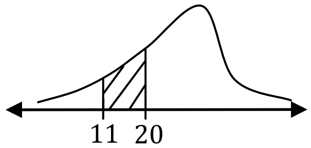
และเราสามารถนำฮิสโทแกรม มาสร้างเป็น “รูปหลายเหลี่ยมของความถี่” โดยลากเส้นเชื่อมตามยอดของแต่ละแท่ง



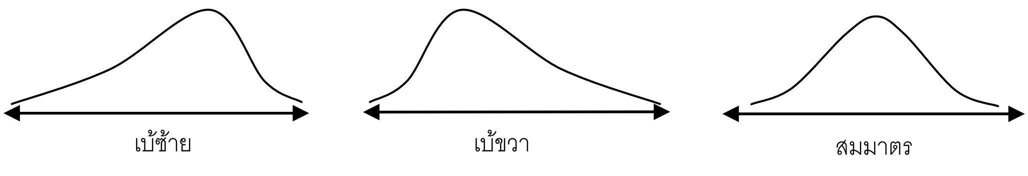
และสุดท้าย เราสามารถนำรูปหลายเหลี่ยมของความถี่ มาปรับให้เป็นเส้นโค้ง ที่เรียกว่า “เส้นโค้งของความถี่” ได้ ดังนี้



ในโค้งความถี่ มักจะไม่นิยมแสดง แกนจำนวนนักเรียน
โดยในโค้งความถี่ เราจะใช้ “พื้นที่ใต้โค้ง” ในการบอกจำนวนข้อมูล
เช่น จำนวนนักเรียนที่ได้คะแนน 11 - 20 จะเป็นสัดส่วนกับพื้นที่ใต้โค้ง จาก 11 ถึง 20

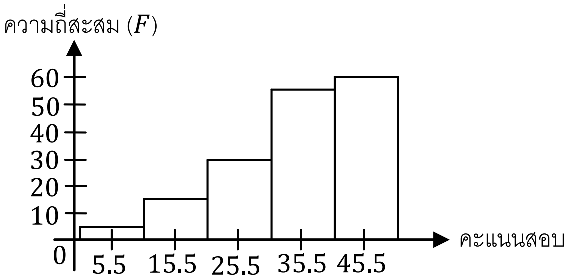


และเราสามารถใช้คำว่า “เบ้” มาบอกลักษณะการเอียงของโค้งความถี่ได้อีกด้วย



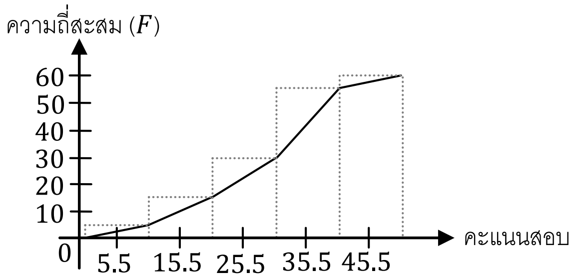
นอกจากนี้ เรายังสามารถเอา “ความถี่สะสม” มาสร้างกราฟคล้ายๆฮิสโทแกรมได้ ดังนี้

คะแนนสอบ	จำนวนนักเรียน (คน)	ความถี่สะสม
1 - 10	3	3
11 - 20	12	15
21 - 30	15	30
31 - 40	24	54
41 - 50	6	60

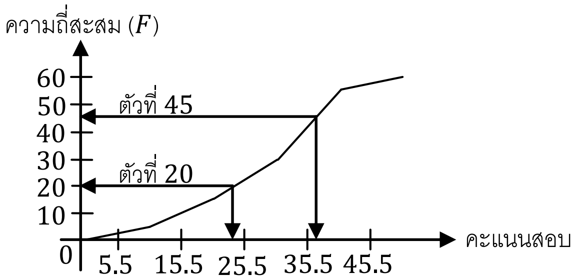


จะเห็นว่ากราฟที่ได้จากความถี่สะสม จะสูงขึ้นเรื่อยๆ (เพราะความถี่สะสมมีแต่บวกเพิ่มไปเรื่อยๆ) และแท่งสุดท้ายจะสูงเท่ากับจำนวนข้อมูล (N)

และสุดท้าย ถ้าลากเส้นเชื่อมตามยอดของแต่ละแท่งของความถี่สะสม ดังรูป เราจะได้กราฟที่เรียกว่า “โอจีฟ”



ประโยชน์ที่สำคัญอย่างหนึ่งของโอจีฟ คือ เราสามารถพล็อตหา ค่าประมาณของข้อมูล ณ ตำแหน่งต่างๆ ตามใจชอบได้ เพราะตำแหน่งข้อมูล จะมีความสัมพันธ์โดยตรง กับ ความถี่สะสม



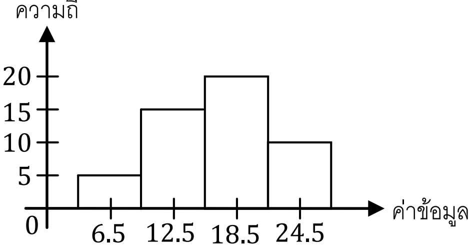
เช่นจากรูป จะได้ว่า

- ข้อมูลตัวที่ 20 มีค่าประมาณ 23 กว่าๆ
- ข้อมูลตัวที่ 45 มีค่าประมาณ 36 กว่าๆ

<u>แบบฝึกหัด</u>

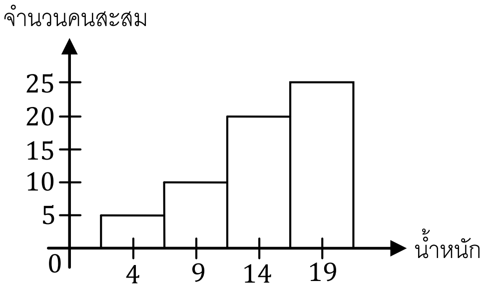
1. จงเติมตารางแจกแจงความถี่ให้สอดคล้องกับแผนภาพต่อไปนี้

1.



ค่าข้อมูล	ความถี่

2.



น้ำหนัก	จำนวนคน

แผนภาพต้น - ใบ

แผนภาพต้นใบ เป็นวิธีนำเสนอข้อมูลอีกแบบหนึ่ง ที่สะดวกในการดูแนวโน้มของค่าข้อมูล
โดยเราจะนำ “หลักหน่วย” ของข้อมูลมาเขียนเป็น “ใบ” และนำ หลักสิบ, ร้อย, พัน, ... มาเขียนเป็น “ลำต้น”

เช่น ข้อมูล

15	24	8	10	32
10	23	18	31	25
9	22	21	37	22

จะนำมาสร้างแผนภาพต้น - ใบ ได้ดังนี้

หลักสิบ	หลักหน่วย					
0	8	9				
1	5	0	0	8		
2	4	3	5	1	2	2
3	2	1	7			

ในแต่ละแถว จะ เรียงลำดับข้อมูล ยังไงก็ได้

และในกรณีที่มีข้อมูล 2 กลุ่ม เราสามารถนำข้อมูลทั้ง 2 กลุ่ม มาเขียนลงใน “ลำต้นเดียวกัน” ได้
เช่น ถ้ามีคะแนนสอบของนักเรียน 2 ห้อง

คะแนนสอบห้อง 5/1

98	102	121	130
105	111	115	123
119	124	120	133

คะแนนสอบห้อง 5/2

121	110	91	115
116	124	114	120
138	124	99	102

จะนำมาเขียนแผนภาพต้น - ใบ โดยใช้ลำต้นเดียวกันได้ ดังนี้

ห้อง 5/1					ห้อง 5/2			
			8	9	1	9		
		5	2	10	2			
	9	5	1	11	0	5	6	4
0	4	3	1	12	1	4	0	4
		3	0	13	8			

<u>แบบฝึกหัด</u>

1. จงเขียนตารางแจกแจงความถี่ จากแผนภาพต้นใบต่อไปนี้

8	6	2	0	3	
9	8	3	1	2	7
10	4	9	3		
11	2	3			

ค่าข้อมูล	ความถี่
80 - 88	
89 - 97	
98 - 106	
107 - 115	

2. ในกรณีที่มีข้อมูลจำนวนมาก การนำเสนอข้อมูลในรูปแบบใดต่อไปนี้ทำให้เห็นการกระจายของข้อมูลได้<u>ชัดเจนน้อยที่สุด</u> [O-NET 52/32]
   1. ตารางแจกแจงความถี่
   2. แผนภาพต้นใบ
   3. ฮิสโทแกรม
   4. การแสดงค่าสังเกตทุกค่า

ค่ากลางข้อมูล

เวลาเรามีข้อมูลหลายๆตัว (เช่น น้ำหนักของเด็ก 30 คน) เรามักจะสนใจว่า “ตรงกลาง” ของข้อมูลชุดนั้นมีค่าเท่าไหร่ และเรามักใช้ค่า “ตรงกลาง” นี้ เป็นตัวแทนของข้อมูลทั้งกลุ่ม เพื่อนำไปวิเคราะห์ หรือ เปรียบเทียบกับข้อมูลชุดอื่น ต่อไป

“ค่ากลางข้อมูล” แปลตรงๆตัวว่า ค่าที่อยู่ตรงกลางของกลุ่มข้อมูล
ค่ากลางข้อมูล มีอยู่หลายชนิด ขึ้นกับว่าเราจะวัดความกลางยังไง
ซึ่งในบทนี้ เราจะได้เรียนค่ากลาง 3 ชนิด ได้แก่

- “ค่าเฉลี่ย” คือ ค่าที่อยู่ตรงกลาง โดยใช้ “ค่าของข้อมูล” เป็นตัววัด
- “มัธยฐาน” คือ ค่าที่อยู่ตรงกลาง โดยใช้ “ตำแหน่งของข้อมูล” เป็นตัววัด
- “ฐานนิยม” คือ ค่าที่ “ซ้ำบ่อยที่สุด” ในกลุ่มข้อมูล

ค่ากลางข้อมูลแต่ละชนิด มีข้อดีข้อเสียต่างกัน ขึ้นกับว่าจะเอาไปใช้ทำอะไร
ข้อมูลบางประเภท ก็ไม่สามารถหาค่ากลางบางชนิดได้
การเลือกค่ากลาง จึงต้องคำนึงถึง ลักษณะข้อมูล และการนำไปใช้

ค่าเฉลี่ย

ค่าเฉลี่ย (Mean) คือ ค่าที่ได้จากการเอาข้อมูลทุกตัวมา "รวม" กัน แล้ว "เฉลี่ย" ด้วยจำนวนตัว
ค่าเฉลี่ย ยังแบ่งต่อไปได้อีกหลายชนิด ขึ้นกับว่าจะ "รวม" และ "เฉลี่ย" ยังไง เช่น

- ค่าเฉลี่ยเลขคณิต (Arithmetric Mean: A.M.) ได้จากการเอาข้อมูลทุกตัวมา "บวก" กัน แล้ว "หาร" ด้วยจำนวนตัว
- ค่าเฉลี่ยเรขาคณิต (Geometric Mean: G.M.) ได้จากการเอาข้อมูลทุกตัวมา "คูณ" กัน แล้ว "ถอดรูท" อันดับที่ตามจำนวนตัว
- ค่าเฉลี่ยฮาร์โมนิค (Harmonic Mean: H.M.) ได้จาก เอาข้อมูลมากลับเศษ - ส่วน แล้วนำไปหาค่าเฉลี่ยเลขคณิต แล้วนำค่าเฉลี่ยเลขคณิตที่ได้ มากลับเศษ - ส่วน กลับไปเป็นแบบเก่า

<u>ตัวอย่าง</u> ข้อมูลชุดหนึ่ง มี 4 จำนวน ประกอบด้วย $2\ ,\ 4\ ,\ 9\ ,\ 18$ จงหาค่าเฉลี่ยเลขคณิต ค่าเฉลี่ยเรขาคณิต และค่าเฉลี่ยฮาร์โมนิค ของข้อมูลชุดนี้

<u>วิธีทำ</u> ค่าเฉลี่ยเลขคณิต $= \frac{2+4+9+18}{4} = \frac{33}{4} = 8.25$

ค่าเฉลี่ยเรขาคณิต $= \sqrt[4]{2 \times 4 \times 9 \times 18} = 6$

ค่าเฉลี่ยฮาร์โมนิค $\rightarrow$ เอาข้อมูลมากลับเศษ - ส่วน ได้เป็น $\frac{1}{2}, \frac{1}{4}, \frac{1}{9}, \frac{1}{18}$

$\rightarrow$ หาค่าเฉลี่ยเลขคณิต ได้ $\frac{\frac{1}{2}+\frac{1}{4}+\frac{1}{9}+\frac{1}{18}}{4} = \frac{\frac{18+9+4+2}{36}}{4} = \frac{33}{36} \times \frac{1}{4} = \frac{11}{48}$

$\rightarrow$ กลับเศษ - ส่วน อีกรอบ แล้วตอบ จะได้ $\frac{48}{11}$ #

ค่าเฉลี่ยเลขคณิต จะเป็นค่ากลางที่ใช้บ่อยกว่าค่ากลางอื่น
ดังนั้น ถ้าได้ยินคำว่า "ค่าเฉลี่ย" เฉยๆ ก็ให้หมายถึง "ค่าเฉลี่ยเลขคณิต" โดยอัตโนมัติ

ในเรื่องสถิตินี้ อีกเดี๋ยวจะมีสูตรมากมาย โผล่ออกมาให้เราทำความเข้าใจ และ ท่อง (ถ้าไม่เข้าใจ)
ดังนั้น เราควรทำความคุ้นเคยกับสัญลักษณ์ที่นิยมใช้ในเรื่องนี้กันก่อน
ในเรื่องสถิติ เรานิยมใช้ N แทน "จำนวนข้อมูล" และใช้ x แทน "ค่าของข้อมูล"

โดยเรานิยมให้ x_1 แทนค่าของข้อมูลตัวที่ 1

x_2 แทนค่าของข้อมูลตัวที่ 2

$\vdots$

x_N แทนค่าของข้อมูลตัวที่ N

เช่น ถ้าข้อมูลชุดหนึ่ง ประกอบด้วย $2\ ,\ 6\ ,\ 9\ ,\ 18$

จะได้ว่า $N = 4$ และ $x_1 = 2$, $x_2 = 6$, $x_3 = 9$, $x_4 = 18$

เราจะใช้สัญลักษณ์ $\bar{x}$ แทนค่าเฉลี่ยเลขคณิต

ดังนั้น เราจะเขียนสูตรได้เป็น $\bar{x} = \frac{x_1+x_2+\ldots+x_N}{N}$

อย่างไรก็ตาม เรานิยมใช้สัญลักษณ์ $\sum$ มาช่วยเขียนสูตรให้กระชับขึ้นได้

โดยสัญลักษณ์ $\sum_{i=1}^{N} x_i$ = ผลบวกของ x_i ทั้งหลาย โดยเริ่มตั้งแต่ $i = 1$ จนถึง $i = N$"

$= x_1 + x_2 + x_3 + \ldots + x_N$ นั่นเอง

ดังนั้น สูตรค่าเฉลี่ยเลขคณิต จะนิยมเขียนเป็น $\bar{x} = \frac{\sum_{i=1}^{N} x_i}{N}$ หรือเขียนแบบย่อๆได้ว่า $\bar{x} = \frac{\sum x_i}{N}$

สูตรนี้ เป็นความสัมพันธ์ของปริมาณ 3 อย่าง คือ "ค่าเฉลี่ย" "ผลรวมข้อมูล" และ "จำนวนข้อมูล"

ถ้าโจทย์บอก 2 อย่างใดๆมา เราต้องสามารถหาตัวที่เหลือได้ทันที ดังนี้

$\boxed{\bar{x} = \frac{\sum x_i}{N}}$ $\boxed{N = \frac{\sum x_i}{\bar{x}}}$ $\boxed{\sum x_i = N \cdot \bar{x}}$ ***

ตัวอย่าง นักเรียนห้องหนึ่งมี 40 คน มีค่าเฉลี่ยของคะแนนสอบเท่ากับ 7.8 คะแนน ต่อมาพบว่าตรวจคะแนนผิด โดยจะต้องเพิ่มคะแนนให้นักเรียนคนหนึ่ง 5 คะแนน และลดคะแนนของนักเรียนอีกคนหนึ่ง 1 คะแนน จงหาค่าเฉลี่ยหลังจากที่แก้ไขคะแนนแล้ว

วิธีทำ ตอนแรก $\bar{x} = 7.8$, $N = 40$ ดังนั้น ผลรวมคะแนน $= 7.8 \times 40 = 312$

แต่ที่ถูก ต้องมีคะแนนค่าหนึ่งเพิ่มขึ้น 5 และอีกค่าหนึ่งลดลง 1

ดังนั้น ผลรวมคะแนนที่ถูกต้อง คือ $312 + 5 - 1 = 316$

ดังนั้น $\bar{x}$ หลังแก้คะแนน $= \frac{316}{40} = 7.9$ คะแนน #

แบบฝึกหัด

1. จงหาค่าเฉลี่ยเลขคณิตของข้อมูล $5 , 8 , 12 , 12 , 16$

2. นักเรียน 8 คน มีค่าเฉลี่ยเลขคณิตของคะแนนสอบคือ 7.5

   ถ้าคะแนนของนักเรียน 7 คนแรก คือ $5 , 8 , 7.5 , 8 , 6 , 7 , 9$ แล้ว จงหาคะแนนของอีกหนึ่งคนที่เหลือ

3. นักเรียนห้องหนึ่ง มี 20 คน มีคะแนนเฉลี่ย คือ 8 แต่ต่อมาพบว่าค่าเฉลี่ยนี้ ไม่ถูกต้อง เนื่องจากอ่านคะแนนของนักเรียนคนหนึ่งผิดไป จากคะแนนที่ถูกคือ 6 คะแนน อ่านผิดเป็น 8 คะแนน จงหาคะแนนเฉลี่ยที่ถูกต้อง

4. นักเรียนห้องหนึ่ง มี 20 คน เป็นผู้หญิง 12 คน ถ้าคะแนนเฉลี่ยของนักเรียนหญิง คือ 8 คะแนน และคะแนนเฉลี่ยของนักเรียนชาย คือ 6 คะแนน จงหาคะแนนเฉลี่ยของนักเรียนทั้งห้อง

5. นักเรียนห้องหนึ่ง มีน้ำหนักเฉลี่ย 50 กก. ถ้าน้ำหนักเฉลี่ยของนักเรียนหญิงคือ 45 กก. และน้ำหนักเฉลี่ยของนักเรียนชายคือ 60 กก. จงหาอัตราส่วนของจำนวนนักเรียนหญิง ต่อ จำนวนนักเรียนชาย

6. นักเรียนห้องหนึ่ง มีอัตราส่วนนักเรียนชาย ต่อ นักเรียนหญิง คือ 2 : 3 ถ้าน้ำหนักเฉลี่ยของนักเรียนชายคือ 65 กก. และน้ำหนักเฉลี่ยของนักเรียนหญิงคือ 50 กก. จงหาน้ำหนักเฉลี่ยของนักเรียนทั้งห้อง

7. สัตว์กลุ่มหนึ่ง มี 30 ตัว ประกอบด้วย วัว 8 ตัว ช้าง 12 ตัว และ ม้า 10 ตัว พบว่าอายุเฉลี่ยของสัตว์กลุ่มนี้ คือ 20 ปี และ อายุเฉลี่ยของวัว เท่ากับ 10 ปี ถ้าอายุรวมของช้าง คือ 300 ปี จงหาอายุเฉลี่ยของม้า

8. ค่าเฉลี่ยเลขคณิตของน้ำหนักของพนักงานของบริษัทหนึ่ง เท่ากับ 48.01 กิโลกรัม บริษัทนี้มีพนักงานชาย 43 คน และพนักงานหญิง 57 คน ถ้าค่าเฉลี่ยเลขคณิตของน้ำหนักพนักงานหญิงเท่ากับ 45 กิโลกรัม แล้ว น้ำหนักของพนักงานชายทั้งหมดรวมกันเท่ากับกี่กิโลกรัม [O-NET 53/28]

9. อายุเฉลี่ยของคนกลุ่มหนึ่งเท่ากับ 31 ปี ถ้าอายุเฉลี่ยของผู้หญิงในกลุ่มนี้เท่ากับ 35 ปี และอายุเฉลี่ยของผู้ชายกลุ่มนี้เท่ากับ 25 ปี แล้ว อัตราส่วนระหว่างจำนวนผู้หญิงต่อจำนวนผู้ชายในกลุ่มเท่ากับเท่าใด [O-NET 50/17]

10. ในการสำรวจน้ำหนักตัวของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ของโรงเรียนแห่งหนึ่ง ซึ่งมี 3 ห้อง มีจำนวนนักเรียน 44, 46 และ 42 คน ตามลำดับ ปรากฏว่ามีค่าเฉลี่ยเลขคณิตเท่ากับ 50 กิโลกรัม แต่พบว่าเครื่องชั่งที่ใช้สำหรับนักเรียนห้องแรกมีความคลาดเคลื่อนทำให้ชั่งน้ำหนักได้ตัวเลขสูงเกินจริงคนละ 1 กิโลกรัม ดังนั้นค่าเฉลี่ยเลขคณิตที่ถูกต้องของน้ำหนักตัวของนักเรียนชั้นมัธยมศึกษาปีที่ 6 นี้เท่ากับกี่กิโลกรัม [O-NET 56/32]

11. ในการแข่งขันกีฬามหาวิทยาลัยโลกครั้งที่ 24 ซึ่งประเทศไทยเป็นเจ้าภาพ มีการส่งรายชื่อนักกีฬาจากประเทศไทย 379 คน มีอายุเฉลี่ย 22 ปี ถ้ามีการถอนตัวนักกีฬาไทยออก 4 คน ซึ่งมีอายุ 24, 25, 25 และ 27 ปี และมีการเพิ่มนักกีฬาไทยอีก 5 คน ซึ่งมีอายุเฉลี่ย 17 ปีแล้ว อายุเฉลี่ยของนักกีฬาจากประเทศไทยจะเท่ากับกี่ปี
[O-NET 51/38]

12. ชายคนหนึ่งตักปลาที่เลี้ยงไว้ในกระชังเพื่อส่งขายจำนวน 500 ตัว ซึ่งมีน้ำหนักโดยเฉลี่ยตัวละ 700 กรัม ในจำนวนนี้เป็นปลาจากกระชังที่หนึ่ง 300 ตัว และจากกระชังที่สอง 200 ตัว ถ้าปลาในกระชังที่หนึ่งมีน้ำหนักเฉลี่ยต่อตัวมากกว่าในกระชังที่สอง 50 กรัม แล้วเขาตักปลาจากกระชังที่สองมากี่กิโลกรัม [O-NET 54/38]

13. ถ้าในปี พ.ศ. 2547 ค่าเฉลี่ยเลขคณิตของอายุพนักงานของบริษัทแห่งหนึ่งเท่ากับ 23 ปี ในปีต่อมา บริษัทได้รับพนักงานเพิ่มขึ้นอีก 20 คน ทำให้ค่าเฉลี่ยเลขคณิตของอายุพนักงานในปี พ.ศ. 2548 เท่ากับ 25 ปี และผลรวมของอายุของพนักงานเพิ่มขึ้นจากปี พ.ศ. 2547 อีก 652 ปี เมื่อสิ้นปี พ.ศ. 2548 บริษัทแห่งนี้มีพนักงานทั้งหมดกี่คน
[O-NET 49/1-29]

ค่าเฉลี่ยถ่วงน้ำหนัก

ในการหาค่าเฉลี่ย เราสามารถ "ถ่วงน้ำหนัก" ให้กับข้อมูลที่นำมาหาค่าเฉลี่ยได้ด้วย
ในกรณีที่ข้อมูลแต่ละตัว มีความสำคัญไม่เท่ากัน เราสามารถกำหนดน้ำหนัก (w) ให้กับข้อมูลแต่ละตัว
ซึ่งจะส่งผลให้ค่าเฉลี่ยเอียงไปตามข้อมูลที่สำคัญ มากกว่า ข้อมูลที่ไม่สำคัญ
เช่น ในการคิดเกรดเฉลี่ย เรามักให้น้ำหนักเกรดบางวิชา มากกว่า อีกวิชา

สูตรของค่าเฉลี่ยถ่วงน้ำหนัก จะหาได้จากสูตร

$$\text{ค่าเฉลี่ยถ่วงน้ำหนัก} = \frac{\sum w_i x_i}{\sum w_i}$$

เอาข้อมูลมาคูณกับน้ำหนักของมันก่อน ค่อยเอามาบวกกัน

หารด้วยผลบวกของน้ำหนักทั้งหมด

<u>ตัวอย่าง</u> จากตารางต่อไปนี้ จงหาเกรดเฉลี่ยแบบไม่ถ่วงน้ำหนัก และเกรดเฉลี่ยแบบถ่วงน้ำหนักด้วยหน่วยกิต

	คณิตศาสตร์	วิทยาศาสตร์	ภาษาไทย	ภาษาฝรั่งเศส	สุขศึกษา
หน่วยกิต	2.5	2	2	1	0.5
เกรด	4	3	3	1	2

<u>วิธีทำ</u> เกรดเฉลี่ยแบบไม่ถ่วงน้ำหนัก $= \frac{4+3+3+1+2}{5} = \frac{13}{5} = 2.6$

เกรดเฉลี่ยแบบถ่วงน้ำหนักด้วยหน่วยกิต $= \frac{(4\times2.5)+(3\times2)+(3\times2)+(1\times1)+(2\times0.5)}{2.5+2+2+1+0.5} = \frac{24}{8} = 3$

จะเห็นว่าเกรดเฉลี่ยแบบถ่วงน้ำหนัก จะเอียงไปทางวิชาที่หน่วยกิตมาก (คณิตศาสตร์, วิทยาศาสตร์, ภาษาไทย) มากกว่าวิชาที่หน่วยกิตน้อย (ภาษาฝรั่งเศส, สุขศึกษา)

(หมายเหตุ: ปกติเวลาโรงเรียนคิดเกรดเฉลี่ย จะคิดด้วยวิธีถ่วงน้ำหนักด้วยหน่วยกิตเสมอ) #

<u>ตัวอย่าง</u> นักเรียนห้องหนึ่ง มี 30 คน เป็นชาย 12 คน และเป็นหญิง 18 คน ถ้าน้ำหนักเฉลี่ยของนักเรียนชายคือ 60 กก. และน้ำหนักเฉลี่ยของนักเรียนหญิงคือ 50 กก. แล้ว จงหาน้ำหนักเฉลี่ยของนักเรียนทั้งห้อง

<u>วิธีทำ</u> ข้อนี้คิดได้หลายวิธี ที่ต้องระวังคือ คำตอบจะ<u>ไม่</u>เท่ากับ $\frac{\text{น้ำหนักเฉลี่ยชาย} + \text{น้ำหนักเฉลี่ยหญิง}}{2}$

เพราะนักเรียนหญิง มีจำนวนมากกว่า ดังนั้น น้ำหนักเฉลี่ยของทั้งห้อง ควรจะเอนไปทางนักเรียนหญิงมากกว่า

นั่นคือ คำตอบ จะเท่ากับ ค่าเฉลี่ยที่ถ่วงน้ำหนัก ด้วย "จำนวนนักเรียน ชาย / หญิง"

ดังนั้น น้ำหนักเฉลี่ยของทั้งห้อง $= \frac{(60\times12)+(50\times18)}{12+18}$

$= \frac{720+900}{30}$

$= \frac{1620}{30} = 54$ ก.ก. #

<u>แบบฝึกหัด</u>

1. จงหาเกรดเฉลี่ย

วิชา	คณิตศาสตร์	ภาษาไทย	คอมพิวเตอร์	พละ
หน่วยกิต	2.5	2	1	0.5
เกรด	2	3.5	4	4

2. ถ้าเกรดเฉลี่ยเท่ากับ 3.0 จงหาหน่วยกิตของวิชา ภาษาจีน

วิชา	ภาษาไทย	ภาษาอังกฤษ	ภาษาจีน	ภาษาละติน
หน่วยกิต	2	2	?	1
เกรด	2	3.5	4	3

3. ผลการสอบวิชาคณิตศาสตร์ของนายคณิต ในชั้นมัธยมศึกษาปีที่ 4 เป็นดังนี้

รหัสวิชา	ค41101	ค42101	ค41102	ค41202
จำนวนหน่วยกิต	1	1.5	1	1.5
เกรด	2.5	3	3.5	2

เกรดเฉลี่ยของวิชาคณิตศาสตร์ของนายคณิต ในชั้นมัธยมศึกษาปีที่ 4 เท่ากับเท่าใด [O-NET 50/16]

4. บริษัทหนึ่งมียอดขายในแต่ละไตรมาสของปี 2557 เป็นตามลำดับดังนี้

17 21 19 23 (หน่วย : ล้านบาท)

การพยากรณ์ยอดขายในไตรมาสถัดไปจะใช้ค่าเฉลี่ยเลขคณิตถ่วงน้ำหนัก ถ้าบริษัทถ่วงน้ำหนักข้อมูลด้วย 1, 1, 1 และ 3 ตามลำดับ แล้ว ค่าเฉลี่ยเลขคณิตถ่วงน้ำหนักของข้อมูลชุดนี้เท่ากับกี่ล้านบาท [O-NET 59/27]

ค่าเฉลี่ย (อันตรภาคชั้นเป็นช่วง)

ในกรณีที่ข้อมูลมาในรูปตารางแจกแจงความถี่ที่มีอันตรภาคชั้นเป็นช่วง เราจะสูญเสียความแม่นยำของข้อมูลไปนิดหน่อย กล่าวคือ เราจะไม่รู้ว่า นักเรียน 3 คน ในชั้น 1-10 ได้คะแนนแบบเป๊ะๆเท่าไหร่

คะแนนสอบ	จำนวนนักเรียน (f)
1 - 10	3
11 - 20	12
21 - 30	15
31 - 40	24
41 - 50	6

ในกรณีนี้ ถ้าเราอยากหา $\bar{x}$ เราจะสมมติว่า "แต่ละตัวในชั้น มีค่าเท่ากับจุดกึ่งกลางชั้น"
เช่น เราจะสมมติว่านักเรียน 3 คน ในชั้น 1-10 ได้คะแนน 5.5

ดังนั้น ผลรวมคะแนนของชั้นแรก $= 3 \times 5.5 = 16.5$
ผลรวมคะแนนของชั้นที่สอง $= 12 \times 15.5 = 186.0$
$\vdots$

จากนั้น เอาคะแนนของทุกชั้นมารวมกัน แล้วหารด้วยจำนวนนักเรียนทั้งหมด ก็จะได้ $\bar{x}$

คะแนนสอบ	จำนวนนักเรียน (f)	จุดกึ่งกลางชั้น (x)	ผลรวมคะแนน (fx)
1 - 10	3	5.5	$3 \times 5.5 = 16.5$
11 - 20	12	15.5	$12 \times 15.5 = 186.0$
21 - 30	15	25.5	$15 \times 25.5 = 382.5$
31 - 40	24	35.5	$24 \times 35.5 = 852.0$
41 - 50	6	45.5	$6 \times 45.5 = 273.0$
	$N = \sum f_i = 60$		$\sum f_i x_i = 1710.0$

ดังนั้น $\bar{x} = \frac{\sum f_i x_i}{N} = \frac{1710}{60} = 28.5$ คะแนน

อย่างไรก็ตาม ถ้าในตารางมี "อันตรภาคชั้นเปิด" เราจะหาค่าเฉลี่ยเลขคณิต<u>ไม่ได้</u>

คะแนนสอบ	จำนวนนักเรียน (f)	จุดกึ่งกลางชั้น
1 - 10	3	5.5
11 - 20	12	15.5
21 - 30	15	25.5
31 - 40	24	35.5
≥ 41	6	???

อันตรภาคชั้นเปิด

หาจุดกึ่งกลางชั้น
ไม่ได้ จึงหา $\bar{x}$ ไม่ได้

ในตารางแจกแจงความถี่ จะมีชั้นเปิดได้ 2 ที่ คือที่ชั้นมากสุด (เช่น > 41) กับที่ชั้นน้อยสุด (เช่น < 10)
ซึ่งในกรณีนี้ เราจะหาค่าเฉลี่ยเลขคณิต<u>ไม่ได้</u> เพราะไม่รู้ว่าจุดกึ่งกลางชั้นเท่ากับเท่าไหร่

แบบฝึกหัด

1. จงหาค่าเฉลี่ยเลขคณิตของข้อมูลต่อไปนี้

คะแนน	ความถี่
1 - 10	2
11 - 20	8
21 - 30	6
31 - 40	4

2. ถ้าค่าเฉลี่ยของคะแนน คือ 9 จงหาค่า x

คะแนน	ความถี่
1 - 4	2
5 - 8	5
9 - 12	x
13 - 16	3

3. ตารางแจกแจงความถี่ แสดงจำนวนนักเรียนในช่วงอายุต่างๆของนักเรียนกลุ่มหนึ่ง เป็นดังนี้

ช่วงอายุ (ปี)	ความถี่ (คน)
1 - 5	4
6 - 10	9
11 - 15	2
16 - 20	5

อายุเฉลี่ยของนักเรียนกลุ่มนี้ เท่ากับกี่ปี [O-NET 50/19]

4. ในการทดสอบความถนัดของนักเรียนกลุ่มหนึ่ง มีตารางแจกแจงความถี่ของผลการสอบดังนี้

ช่วงคะแนน	ความถี่ (คน)
0 - 4	4
5 - 9	5
10 - 14	x
15 - 19	7

ถ้าค่าเฉลี่ยเลขคณิตของคะแนนสอบเท่ากับ 11 แล้ว นักเรียนที่สอบได้คะแนนในช่วง 5 - 14 คะแนนมีจำนวนคิดเป็นร้อยละของนักเรียนกลุ่มนี้เท่ากับเท่าใด [O-NET 51/35]

5. ในการสำรวจน้ำหนักตัว ของนักเรียนในชั้นเรียนที่มีนักเรียน 30 คน เป็นดังนี้

น้ำหนัก (กิโลกรัม)	ความถี่สะสม (คน)
30 – 49	10
50 – 69	26
70 – 89	30

ค่าเฉลี่ยเลขคณิตของน้ำหนักตัวของนักเรียนในชั้นเรียนนี้เท่ากับกี่กิโลกรัม [O-NET 54/36]

การหาค่าเฉลี่ยโดยการลดทอนข้อมูล

สมบัติที่สำคัญอย่างหนึ่งของค่าเฉลี่ย คือ เราสามารถ "แปลง" ข้อมูล กับ ค่าเฉลี่ยด้วยสูตรเดียวกันได้
พูดง่ายๆก็คือ ถ้าเราเอาข้อมูลเก่า มาบวกลบคูณหารด้วยตัวเลขอะไรก็ได้ เป็นข้อมูลชุดใหม่
จะได้ ค่าเฉลี่ยของข้อมูลชุดใหม่ มีค่าเปลี่ยนไปจาก ค่าเฉลี่ยเดิม ด้วยการบวกลบคูณหารแบบเดียวกัน

เช่น ถ้าเรามีข้อมูล x_i ซึ่งประกอบด้วย $2\ ,\ 6\ ,\ 10\ ,\ 18$ มี $\bar{x} = \frac{2+6+10+18}{4} = 9$
ต่อมา เราแปลงข้อมูล x_i ให้เป็นข้อมูล y_i ชุดใหม่โดย การ "คูณ 2 แล้วลบด้วย 8 หารด้วย 4" ดังนี้

x_i	คูณ 2 ลบ 8 หารด้วย 4 $\Rightarrow$	y_i
2	$\rightarrow \frac{(2\times2)-8}{4}$	-1
6	$\rightarrow \frac{(6\times2)-8}{4}$	1
10	$\rightarrow \frac{(10\times2)-8}{4}$	3
18	$\rightarrow \frac{(18\times2)-8}{4}$	7
$\bar{x} = 9$		$\bar{y} = ?$

ถ้าจะหาค่าเฉลี่ยของข้อมูล y_i ชุดใหม่ เราไม่ต้องหาตรงๆแล้ว แต่ให้เอา $\bar{x}$ มา "คูณ 2 ลบ 8 หารด้วย 4" ได้เลย
นั่นคือ จะได้ $\bar{y} = \frac{(9\times2)-8}{4} = 2.5$
หรือถ้าหาแบบตรงๆ ก็จะได้ $\bar{y} = \frac{(-1)+1+3+7}{4} = 2.5$ เท่ากัน

<u>ตัวอย่าง</u> นักเรียนห้องหนึ่ง มีค่าเฉลี่ยของคะแนนวิชาคณิตศาสตร์เท่ากับ 74 คะแนน ต่อมา คุณครูพบว่าต้องบวกคะแนนจิตพิสัยให้ทุกๆคนอีกคนละ 10 คะแนน จงหาค่าเฉลี่ยของคะแนนหลังจากที่บวกคะแนนจิตพิสัยเข้าไปแล้ว
<u>วิธีทำ</u> จากสมบัติของค่าเฉลี่ยเลขคณิต ถ้าข้อมูลทุกตัวเพิ่มขึ้น 10 ก็จะทำให้ค่าเฉลี่ยเพิ่มขึ้น 10 ด้วย
ดังนั้น ค่าเฉลี่ยของคะแนนหลังจากที่บวกคะแนนจิตพิสัยเข้าไปแล้ว จะเท่ากับ $74 + 10 = 84$ คะแนน #

นอกจากนี้ ความรู้ในเรื่องการแปลง ยังช่วยให้เราหลีกเลี่ยงการคำนวณเลขเยอะๆยุ่งๆ ได้
โดยในกรณีที่ข้อมูลมีค่าเยอะ หรือเป็นตัวเลขทศนิยมที่คำนวณลำบาก เราจะมีวิธีหา $\bar{x}$ โดยการ "ลดทอนข้อมูล" ดังนี้

1. แปลงข้อมูลให้เป็น "ตัวเลขง่ายๆ" ก่อน
   สูตรแปลงที่นิยมใช้ คือ "ลบด้วยตัวที่ซ้ำมากสุด แล้วหารทอนด้วยความกว้างชั้น"
2. หา $\bar{x}$ ของตัวเลขง่ายๆ
3. แปลง $\bar{x}$ ในข้อ 2 กลับ ด้วยการย้อนสูตรที่ใช้แปลงในข้อ 1

<u>ตัวอย่าง</u> จงหาค่าเฉลี่ยเลขคณิตของข้อมูล 9.87 , 12.87 , 12.87 , 16.87 , 18.87

<u>วิธีทำ</u> ถ้าคิดแบบตรงไปตรงมา จะได้ $\bar{x} = \frac{9.87+12.87+12.87+16.87+18.87}{5} = 14.27$

อีกวิธี คือ เราจะลดทอนข้อมูลโดยเอาข้อมูลทุกตัวมาลบด้วย 12.87 เพื่อให้ ".87" หายไป

$$\begin{array}{ccc} 9.87 & & -3 \\ 12.87 & -12.87 & 0 \\ 12.87 & \Longrightarrow & 0 \\ 16.87 & & 4 \\ 18.87 & & 6 \end{array}$$

จะได้ข้อมูลชุดใหม่แบบตัวเลขง่ายๆ คือ $-3\ ,\ 0\ ,\ 0\ ,\ 4\ ,\ 6\ \ \rightarrow$ ค่าเฉลี่ย $= \frac{-3+0+0+4+6}{5} = \frac{7}{5} = 1.4$

$$??? \xleftarrow{+12.87} 1.4$$

เอา 1.4 มาย้อนสูตรแปลงกลับ โดยการบวก 12.87 จะได้ $\bar{x} = 14.27$ #

<u>ตัวอย่าง</u> จากตารางแสดงคะแนนสอบของนักเรียน 40 คน จงหาค่าเฉลี่ยเลขคณิต

คะแนนสอบ	จำนวนนักเรียน
10 - 19.9	7
20 - 29.9	11
30 - 39.9	17
40 - 49.9	5

<u>วิธีทำ</u> ในกรณีนี้ ถ้าจะหา $\bar{x}$ เราจะสมมุติว่าทุกตัวในชั้น มีค่าประมาณจุดกึ่งกลางชั้น

ถ้ายอมเหนื่อยหน่อย ก็ทำแบบตรงไปตรงมาดังนี้

คะแนนสอบ	จำนวนนักเรียน (f)	จุดกึ่งกลางชั้น (x)	รวมคะแนนในชั้น (fx)
10 - 19.9	7	14.95	104.65
20 - 29.9	11	24.95	274.45
30 - 39.9	17	34.95	594.15
40 - 49.9	5	44.95	224.75
	40		1198.00

ซึ่งจะได้ $\bar{x} = \frac{1198}{40} = 29.95$

แต่ถ้าไม่อยากเหนื่อย ก็มีอีกมีวิธี คือ เราจะแปลงจุดกึ่งกลางชั้นให้เป็นตัวเลขง่ายๆก่อน

วิธีแปลงที่นิยมคือ ลบด้วยตัวที่ซ้ำมากสุด $(= 34.95)$ แล้วหารทอนด้วยความกว้างชั้น $(= 10)$

จากนั้น ลืมจุดกึ่งกลางชั้นไปซะ แล้วหาค่าเฉลี่ยของผลแปลงแทน ดังนี้

คะแนนสอบ	จำนวนนักเรียน (f)	~~จุดกึ่งกลางชั้น (x)~~	ผลแปลง $\left(d = \frac{x-34.95}{10}\right)$	รวมคะแนนในชั้น (fd)
10 - 19.9	7	~~14.95~~	-2	-14
20 - 29.9	11	~~24.95~~	-1	-11
30 - 39.9	17	~~34.95~~	0	0
40 - 49.9	5	~~44.95~~	1	5
	40			-20

จะได้ค่าเฉลี่ยของผลแปลง คือ $\bar{d} = \frac{-20}{40} = -0.5$

จากนั้น แปลง -0.5 กลับ ให้เป็นค่าเฉลี่ยของข้อมูลเดิม โดยย้อนสูตรที่ใช้แปลง

$$\begin{tikzcd} x \arrow[r, bend left, "-34.95\ ,\ \div 10"] & d \arrow[l, bend left, "\times 10\ ,\ +34.95"] \end{tikzcd}$$

ขามา เราลบ 34.95 แล้วหารด้วย 10

ดังนั้น ขากลับ ต้องคูณ 10 แล้วบวก 34.95

จะได้ $\bar{x} = (-0.5\times 10) + 34.95 = 29.95$ #

<u>แบบฝึกหัด</u>

1. คุณครูที่โรงเรียนแห่งหนึ่ง มีอายุเฉลี่ย 35 ปี หลังจากนั้น 5 ปี มีครูใหม่เข้ามาเพิ่ม 10 คน ทำให้อายุเฉลี่ยเปลี่ยนเป็น 36 ปี ถ้าครูใหม่ทั้ง 10 คน มีอายุ 28 ปี ทุกคน จงหาว่าเดิม โรงเรียนแห่งนี้มีครูกี่คน

2. ในการสอบย่อยเพื่อเก็บคะแนน พบว่าคะแนนดิบของนักเรียน 5 คน คือ $26 , 29 , 25 , 28 , 27$ อย่างไรก็ตาม คะแนนดิบนี้ ต้องนำไปคูณ $\frac{2}{3}$ ก่อน จึงจะได้คะแนนเก็บจริงๆ จงหาค่าเฉลี่ยของคะแนนเก็บจริงๆ ของนักเรียนทั้ง 5 คนนี้

3. จงหาค่าเฉลี่ยเลขคณิตโดยวิธีลดทอนข้อมูล

ส่วนสูง	จำนวนนักเรียน (f)			
140 - 149	3			
150 - 159	12			
160 - 169	18			
170 - 179	7			

4. คะแนนสอบวิชาคณิตศาสตร์ของนักเรียน 50 คน มีตารางแจกแจงความถี่ดังนี้

ช่วงคะแนน	จำนวนนักเรียน (คน)
1 – 20	3
21 – 40	5
41 – 60	13
61 – 80	20
81 – 100	9

ค่าเฉลี่ยเลขคณิตของคะแนนสอบนี้เท่ากับเท่าใด [O-NET 56/40]

5. คะแนนสอบวิชาคณิตศาสตร์ของนักเรียนกลุ่มหนึ่ง เป็นดังตารางแจกแจงความถี่

คะแนน	ความถี่
20 – 29	7
30 – 39	10
40 – 49	6
50 – 59	7
60 – 69	6
70 – 79	8
80 – 89	6

ค่าเฉลี่ยของคะแนนสอบนี้เป็นเท่าใด [O-NET 58/30]

6. ความสัมพันธ์ระหว่างกำไร (y) และราคาทุน (x) ของสินค้าในร้านแห่งหนึ่งเป็นไปตามสมการ $y = 2x - 30$ ถ้าราคาทุนของสินค้า 5 ชนิด คือ 31, 34, 35, 36 และ 39 บาท แล้ว ค่าเฉลี่ยเลขคณิตของกำไรในการขายสินค้า 5 ชนิดนี้ เท่ากับกี่บาท [O-NET 50/18]

สมบัติของค่าเฉลี่ย

ค่าเฉลี่ยเลขคณิต มีสมบัติพิเศษ อยู่หลายอย่าง

อย่างแรก $\bar{x}$ จะมี “ค่า” อยู่ตรงกลาง แต่ไม่แน่ว่าจะมี “ตำแหน่ง” อยู่ตรงกลางหรือเปล่า

เช่น ข้อมูล $2\,,\,3\,,\,5\,,\,7\,,\,13\,,\,24$ $\rightarrow$ จะได้ $\bar{x} = \frac{2+3+5+7+13+24}{6} = 9$

จะเห็นว่า มีข้อมูล 4 ตัว $(= 2, 3, 5, 7)$ ที่น้อยกว่า 9 แต่มีข้อมูล 2 ตัว $(= 13, 24)$ ที่มากกว่า 9

จะเห็นว่า $\bar{x}$ ไม่ได้มี “ตำแหน่ง” อยู่ตรงกลาง

อย่างไรก็ตาม $\bar{x}$ จะมี “ค่า” อยู่ตรงกลาง

ถ้าเราเทียบ $\bar{x}$ กับข้อมูลแต่ละตัว จะเห็นว่ามีข้อมูล “บางตัว” มากกว่า $\bar{x}$ และมีข้อมูล “บางตัว” น้อยกว่า $\bar{x}$

แต่ถ้าหักลบกันออกมาแล้ว ส่วนที่มากกว่า $\bar{x}$ กับ ส่วนที่น้อยกว่า $\bar{x}$ จะหักกันหมดพอดี

เช่น ข้อมูล $2\,,\,3\,,\,5\,,\,7\,,\,13\,,\,24$ $\rightarrow$ จะได้ $\bar{x} = \frac{2+3+5+7+13+24}{6} = 9$

2 น้อยกว่า 9 อยู่	7	13 มากกว่า 9 อยู่	4
3 น้อยกว่า 9 อยู่	6	24 มากกว่า 9 อยู่	15
5 น้อยกว่า 9 อยู่	4		
7 น้อยกว่า 9 อยู่	2		
	$= 19$		$= 19$

หรือพูดอีกแบบ ถ้าเราพิจารณาค่าของ $x_i - \bar{x}$ จะเห็นว่า x_i ที่มากกว่า $\bar{x}$ จะทำให้ $x_i - \bar{x}$ เป็นบวก

x_i ที่น้อยกว่า $\bar{x}$ จะทำให้ $x_i - \bar{x}$ เป็นลบ

แต่ถ้าเรานำ $x_i - \bar{x}$ ทั้งหมด มาบวกกัน จะได้ผลหักล้างกันเป็น 0 เสมอ

สมบัติข้อนี้ เขียนเป็นประโยคสัญลักษณ์ได้ว่า $\sum(x_i - \bar{x}) = 0$

เช่น จากตัวอย่างที่แล้ว
$$\begin{aligned}\sum(x_i - \bar{x}) &= (2-9) + (3-9) + (5-9) + (7-9) + (13-9) + (24-9)\\ &= (-7) + (-6) + (-4) + (-2) + 4 + 15\\ &= 0\end{aligned}$$

สมบัติอย่างที่สอง คือ $\sum(x_i - a)^2$ จะมีค่าน้อยที่สุด เมื่อ $a = \bar{x}$

ข้อนี้ ต่างจากสมบัติข้อแรก ตรงที่มีการยกกำลังสอง ซึ่งจะทำให้เลขลบกลายเป็นบวก

เช่น จากตัวอย่างที่แล้ว ถ้าแทน a ด้วย $\bar{x}$ จะได้

$$\begin{aligned}\sum(x_i - \bar{x})^2 &= (2-9)^2 + (3-9)^2 + (5-9)^2 + (7-9)^2 + (13-9)^2 + (24-9)^2\\ &= (-7)^2 + (-6)^2 + (-4)^2 + (-2)^2 + 4^2 + 15^2\\ &= 49 + 36 + 16 + 4 + 16 + 225\\ &= 346\end{aligned}$$

จะเห็นว่า คราวนี้ ไม่ได้ 0 เพราะการยกกำลังสอง จะทำให้เลขลบกลายเป็นบวก ตัวเลขจึงไม่หักกันเหมือนสมบัติข้อแรก

แต่ถึงจะไม่ได้ 0 สมบัติข้อนี้ จะรับประกันว่า 346 เป็นค่าที่น้อยที่สุด ในบรรดา $\sum(x_i - a)^2$ ทั้งหมด

พูดง่ายๆก็คือ ถ้า a เป็นค่าอื่นที่ไม่ใช่ $\bar{x}$ จะไม่มีวันได้ $\sum(x_i - a)^2$ น้อยกว่า 346

เช่น ถ้าแทน a ด้วย 10 จะได้

$$\begin{aligned}\sum(x_i-10)^2 &= (2-10)^2+(3-10)^2+(5-10)^2+(7-10)^2+(13-10)^2+(24-10)^2\\ &= (-8)^2+(-7)^2+(-5)^2+(-3)^2+3^2+14^2\\ &= 64+49+25+9+9+196\\ &= 352 \quad \rightarrow \text{มากกว่า } 346\end{aligned}$$

หรือ ถ้าแทน a ด้วย 8.5 จะได้

$$\begin{aligned}\sum(x_i-8.5)^2 &= (2-8.5)^2+(3-8.5)^2+(5-8.5)^2+(7-8.5)^2+(13-8.5)^2+(24-8.5)^2\\ &= (-6.5)^2+(-5.5)^2+(-3.5)^2+(-1.5)^2+(4.5)^2+(15.5)^2\\ &= 42.5+30.25+12.25+2.25+20.25+240.25\\ &= 347.5 \quad \rightarrow \text{มากกว่า } 346\end{aligned}$$

ดังนั้น จะเห็นว่า $\sum(x_i-a)^2$ จะมีค่าน้อยที่สุด เมื่อ $a=\bar{x}$

<u>แบบฝึกหัด</u>

1. ให้ x เป็นจำนวนจริงใดๆ จงหาค่าต่ำสุดของ
   1. $(2-x)^2+(4-x)^2+(9-x)^2$
   2. $(x-2)^2+(x-4)^2+(x-9)^2$
   3. $(x+2)^2+(x+4)^2+(x+9)^2$
   4. $2(2-x)^2+(x+5)^2+(x-1)^2$

มัธยฐาน

ค่ากลางตัวถัดมาที่ต้องเรียน คือ “มัธยฐาน” (Median) หรือเรียกสั้นๆว่า Med
มัธยฐาน เป็นค่ามี “ตำแหน่ง” อยู่ตรงกลาง เมื่อเรียงลำดับข้อมูลจากน้อยไปมาก (หรือมากไปน้อยก็ได้)
หรือพูดง่ายๆก็คือ Med จะเป็นจุดที่แบ่งจำนวนข้อมูลออกเป็น 2 ส่วน เท่าๆกัน

การหา Med ประกอบด้วย 2 ขั้นตอน
ขั้นแรก ต้องหาก่อนว่า “ตำแหน่งตรงกลาง” คือตำแหน่งที่เท่าไหร่
เช่น ถ้ามีข้อมูล 3 ตัว ตำแหน่งตรงกลางคือตัวที่ 2
ถ้ามีข้อมูล 4 ตัว ตำแหน่งตรงกลางคือตัวที่ 2.5 (คืออยู่ระหว่างตัวที่ 2 กับตัวที่ 3)
ถ้ามีข้อมูล 5 ตัว ตำแหน่งตรงกลางคือตัวที่ 3

จะได้สูตร คือ $\boxed{\text{ตำแหน่งมัธยฐาน} = \frac{N+1}{2}}$

ขั้นถัดมา เอาตำแหน่งที่ได้ไปหาค่า
โดยต้องเรียงข้อมูลจากน้อยไปมาก (หรือมากไปน้อยก็ได้) แล้วจิ้มนิ้วไล่นับจนกว่าจะถึงตำแหน่งที่ต้องการ
ถ้าตำแหน่งที่คำนวณได้ ไปตกอยู่ตรงกลางระหว่างข้อมูลสองค่า ให้เอาสองค่านั้นบวกกัน หารด้วยสอง

<u>ตัวอย่าง</u> จงหาค่ามัธยฐานของข้อมูลชุดหนึ่ง ซึ่งประกอบด้วย $5\,,\ 15\,,\ 12\,,\ 21\,,\ 13\,,\ 9\,,\ 18$
<u>วิธีทำ</u> มีข้อมูล 7 ตัว ดังนั้น มัธฐานจะอยู่ตำแหน่งที่ $\frac{7+1}{2} = 4$
เรียงข้อมูลจากน้อยไปมาก จะได้ $5\,,\ 9\,,\ 12\,,\ 13\,,\ 15\,,\ 19\,,\ 21$ จะเห็นว่าตัวที่ 4 มีค่า $= 13$
(จริงๆไม่ต้องเรียงข้อมูลจนหมดทุกตัวก็ได้ แค่เรียงให้ถึงตำแหน่งที่ 4 ก็พอ)
ดังนั้น จะได้ $\text{Med} = 13$ #

<u>ตัวอย่าง</u> จงหาค่ามัธยฐานของข้อมูลชุดหนึ่ง ซึ่งประกอบด้วย $12\,,\ 18\,,\ 22\,,\ 25$
<u>วิธีทำ</u> มีข้อมูล 4 ตัว ดังนั้น มัธฐานจะอยู่ตัวที่ $\frac{4+1}{2} = 2.5$
จะเห็นว่าข้อนี้ใจดี เรียงข้อมูลมาให้แล้ว และตำแหน่งที่ 2.5 จะอยู่ตรงกลางระหว่างตัวที่ 2 กับตัวที่ 3
ถ้าเป็นแบบนี้ ให้เอาตัวที่ 2 กับตัวที่ 3 บวกกัน หารด้วย 2 จะได้ $\text{Med} = \frac{18+22}{2} = 20$ #

ในกรณีที่ข้อมูลมาในรูปตารางแจกแจงความถี่ เราจะหา “ตำแหน่ง” ของข้อมูลยากขึ้นนิดหน่อย
สิ่งที่ต้องรำลึกอยู่เสมอ คือ ตำแหน่งของข้อมูล จะมีความเกี่ยวข้องโดยตรงกับ “ความถี่”

คะแนนสอบ	จำนวนนักเรียน (f)	
14	2	ตัวที่ $1\,,2$
15	3	ตัวที่ $3\,,4\,,5$
16	5	ตัวที่ $6\,,7\,,8\,,9\,,10$
17	3	ตัวที่ $11\,,12\,,13$

จะเห็นว่า “ตำแหน่ง” ของข้อมูล จะได้จากการนำ “ความถี่” มาบวกสะสมไปเรื่อยๆ

บางคนจะนิยมสร้างช่อง “ความถี่สะสม” เพื่อความสะดวกรวดเร็วในการระบุตำแหน่งข้อมูล

คะแนนสอบ	จำนวนนักเรียน (f)	ความถี่สะสม (F)	
14	2	2	ตัวที่ 1 , 2
15	3	5	ตัวที่ 3 , 4 , 5
16	5	10	ตัวที่ 6 , 7 , 8 , 9 , 10
17	3	13	ตัวที่ 11 , 12 , 13

ความถี่สะสม จะบอกว่า “ตัวสุดท้ายของชั้น” คือตัวที่เท่าไหร่

ดังนั้น ถ้าอยากหา “ตัวที่ k” ก็แค่ดูว่าความถี่สะสม “เลย k” ไปในชั้นไหน

เช่น ข้อมูลตัวที่ 4 จะอยู่ในชั้นที่ 2 เพราะ ความถี่สะสม เลย 4 ในชั้นที่ 2 ($F = 5$)

ข้อมูลตัวที่ 8 จะอยู่ในชั้นที่ 3 เพราะ ความถี่สะสม เลย 8 ในชั้นที่ 3 ($F = 10$)

ข้อมูลตัวที่ 10 จะอยู่ในชั้นที่ 3 และเป็นตัวสุดท้ายของชั้นที่ 3

ข้อมูลตัวที่ 11 จะอยู่ในชั้นที่ 4 เพราะ ความถี่สะสม เลย 11 ในชั้นที่ 4 ($F = 13$) เป็นต้น

<u>ตัวอย่าง</u> จงหามัธยฐานของข้อมูลต่อไปนี้

น้ำหนัก	จำนวนนักเรียน (f)
50	6
51	13
52	11
53	5

<u>วิธีทำ</u> อันดับแรก สร้างช่อง “ความถี่สะสม” ก่อน เพื่อความสะดวกในการอ้างตำแหน่ง

น้ำหนัก	จำนวนนักเรียน (f)	ความถี่สะสม (F)
50	6	6
51	13	19
52	11	30
53	5	35

จะเห็นว่ามีข้อมูลทั้งหมด 35 ตัว ดังนั้น Med จะอยู่ตำแหน่งที่ $\frac{35+1}{2} = 18$

ย้ำอีกที ว่า 18 ไม่ใช่ “ค่า” มัธยฐาน แต่เป็น “ตำแหน่ง” มัธยฐาน

ถัดมา หาว่า ตัวที่ 18 มีค่าเท่าไหร่ เราต้องหาก่อน ว่า ตัวที่ 18 อยู่ในชั้นไหน

จะเห็นว่าความถี่สะสม เลย 18 ในชั้นที่ 2 ($F = 19$) ดังนั้น ตัวที่ 18 อยู่ในชั้นที่ 2

ดังนั้น จะได้ $\text{Med} = 51$ #

หมายเหตุ: นักเรียนจำนวนมาก มักสับสนระหว่าง “ตำแหน่ง” กับ “ค่า”

สูตร $\frac{N+1}{2}$ เป็นสูตรที่ใช้หา “ตำแหน่ง” มัธยฐาน เท่านั้น ไม่ใช่ “ค่า” มัธยฐาน

เรายังต้องเอาตำแหน่งดังกล่าว ไปหา “ค่า” ต่ออีก

ตัวอย่าง จงหามัธยฐานของข้อมูลต่อไปนี้

คะแนนสอบ	ความถี่สะสม
6	5
10	12
12	20
15	32
18	40

วิธีทำ ข้อนี้ โจทย์ตั้งใจหลอก โดยให้ “ความถี่สะสม” มา แทนที่จะเป็น “ความถี่” เฉยๆ เหมือนปกติ

ถ้าไม่สังเกตดีๆ เราอาจคิดว่าโจทย์ให้ “ความถี่” มา แล้วดันไปสร้างช่อง “ความถี่สะสม” เพิ่ม จะได้คำตอบที่ผิด

จะเห็นว่าข้อนี้มีข้อมูลทั้งหมด 40 ตัว ดังนั้น Med จะอยู่ตำแหน่งที่ $\frac{40+1}{2} = 20.5$

ย้ำอีกที ว่า 20.5 ไม่ใช่ “ค่า” มัธยฐาน แต่เป็น “ตำแหน่ง” มัธยฐาน

ตัวที่ 20.5 จะอยู่ตรงกลาง ระหว่างตัวที่ 20 กับ ตัวที่ 21 $\rightarrow$ ต้องเอาตัวที่ 20 กับ 21 มาบวกกัน แล้ว $\div 2$

ตัวที่ 20 $\rightarrow$ เป็นตัวสุดท้ายของชั้นที่ 3 พอดี ดังนั้น ตัวที่ 20 มีค่า $= 12$

ตัวที่ 21 $\rightarrow$ ความถี่สะสม เลย 21 ในชั้นที่ 4 $(F = 32)$ ดังนั้น ตัวที่ 21 อยู่ในชั้นที่ 4 มีค่า $= 15$

ดังนั้น จะได้ $\text{Med} = \frac{12+15}{2} = 13.5$ #

แบบฝึกหัด

1. จงหามัธยฐานของข้อมูลต่อไปนี้

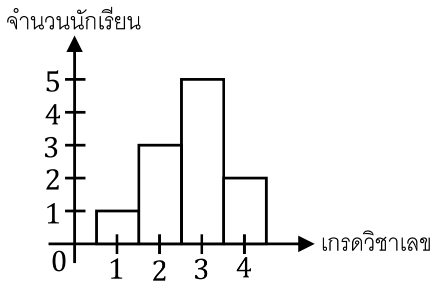
   1. $2, 5, 8, 13, 14, 14, 18, 20$
   2. $5, 2, 8, 10, 7, 11, 9$
   3.

0	|	5	8		
1	|	0	2	2	
2	|	5	6	7	7
3	|	1			

   4.

คะแนน	ความถี่
7	2
8	8
9	12
10	6

5.



6.

คะแนน	ความถี่สะสม
21	2
22	8
23	12
24	15
25	16

7. $2 , 4 , 6 , \ldots , 30$

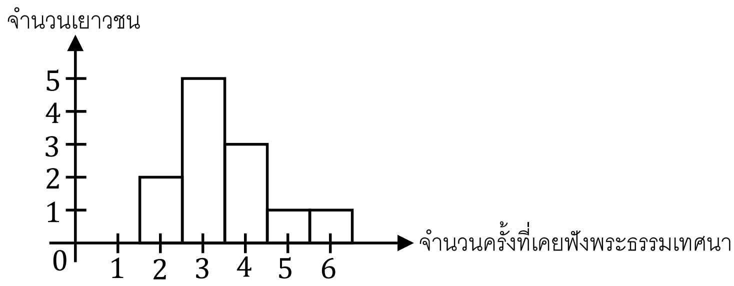
8. $5 , 7 , 9 , 11 , \ldots , 101$

2. จากแผนภาพต้น – ใบของข้อมูลชุดหนึ่งเป็นดังนี้

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

จงหาค่าเฉลี่ยเลขคณิต และ มัธยฐาน ของข้อมูลชุดนี้ [O-NET 57/30]

3. จากการสอบถามเยาวชนจำนวน 12 คน ว่าเคยฟังพระธรรมเทศนามาแล้วจำนวนกี่ครั้ง ปรากฏผลดังแสดงในแผนภาพต่อไปนี้ [O-NET 52/33]



มัธยฐานของข้อมูลชุดนี้เท่ากับกี่ครั้ง

4. กำหนดให้ข้อมูลชุดหนึ่ง คือ $10,\ 3,\ x,\ 6,\ 6$ ถ้าค่าเฉลี่ยเลขคณิตของข้อมูลชุดนี้ มีค่าเท่ากับมัธยฐาน แล้ว x มีค่าเท่ากับเท่าใด [O-NET 50/20]

มัธยฐาน (อันตรภาคชั้นเป็นช่วง)

ในกรณีที่ข้อมูลมาในรูปตารางแจกแจงความถี่ที่มีอันตรภาคชั้นเป็นช่วง เราจะคำนวณ Med ยากขึ้นเยอะเลย

คะแนนสอบ	จำนวนนักเรียน (f)
1 - 10	8
11 - 20	10
21 - 30	20
31 - 40	12

ในกรณีนี้ ถ้าจะหา Med จะมีขั้นตอนดังนี้

1. สร้างช่อง "ความถี่สะสม" เพื่อความสะดวกในการหาค่าข้อมูล ณ ตำแหน่งต่างๆ
2. หา "ตำแหน่ง" มัธยฐาน ด้วยสูตรใหม่

$$\boxed{\text{ตำแหน่งมัธยฐาน} = \frac{N}{2}}$$

สูตรนี้ ใช้กับกรณี อันตรภาคชั้นเป็นช่วงเท่านั้น
ถ้ามาเป็นตัวๆ ให้ใช้สูตร $\frac{N+1}{2}$ เหมือนเก่า

3. ใช้ช่องความถี่สะสม หาว่ามัธยฐาน อยู่ชั้นไหน แล้วใช้สูตรไหนก็ได้ จาก 2 สูตรต่อไปนี้

ความถี่สะสมชั้นที่ต่ำกว่า Med

$$\text{Med} = L + \left(\frac{\frac{N}{2} - F_L}{f_M}\right) I$$

ความกว้างชั้น Med

ขอบล่างชั้น Med

ความถี่ชั้น Med

สูตรนี้ จะเริ่มจากขอบล่าง
แล้วหาว่าต้องเดินหน้าต่อไปอีกเป็นสัดส่วนเท่าไหร่ของชั้น
โดยจะดูว่า Med ล้ำเข้าไปในชั้นนั้นๆ เป็นสัดส่วนเท่าไหร่

$$\text{สัดส่วนที่ล้ำ} = \frac{\text{Med ล้ำไปในชั้น กี่ตัว}}{\text{จำนวนตัวในชั้น}}$$

ความถี่สะสมชั้น Med

$$\text{Med} = U - \left(\frac{F_M - \frac{N}{2}}{f_M}\right) I$$

ความกว้างชั้น Med

ขอบบนชั้น Med

ความถี่ชั้น Med

สูตรนี้ จะเริ่มจากขอบบน
แล้วหาว่าต้องถอยกลับมาเป็นสัดส่วนเท่าไหร่ของชั้น
โดยจะดูว่า Med พร่องไปจากชั้นนั้นๆ เป็นสัดส่วนเท่าไหร่

$$\text{สัดส่วนที่พร่อง} = \frac{\text{ขาดกี่ตัว Med ถึงจะเลื่อนไปเต็มชั้น}}{\text{จำนวนตัวในชั้น}}$$

เช่น

คะแนนสอบ	จำนวนนักเรียน (f)	ความถี่สะสม (F)
1 - 10	8	8
11 - 20	10	18
21 - 30	20	38
31 - 40	12	50

$N = 50$, Med อยู่ตัวที่ $\frac{50}{2} = 25 =$ ชั้นที่ 3

$L = 20.5$, $U = 30.5$, $I = 10$

$F_L = 18$, $F_M = 38$, $f_M = 20$

$$\begin{aligned}\text{Med} &= 20.5 + \left(\frac{25-18}{20}\right)(10) \\ &= 20.5 + 3.5 \\ &= 24\end{aligned}$$

หรือถ้าใช้อีกสูตร

$$\begin{aligned}\text{Med} &= 30.5 - \left(\frac{38-25}{20}\right)(10) \\ &= 30.5 - 6.5 \\ &= 24\end{aligned}$$

เท่ากัน

ตัวอย่างอีกอัน เช่น

คะแนนสอบ	จำนวนนักเรียน (f)	ความถี่สะสม (F)
1 - 12	8	8
13 - 24	12	20
25 - 36	10	30
37 - 48	7	37
49 - 60	3	40

$N = 40$

Med อยู่ตัวที่ $\frac{40}{2} = 20 =$ ตัวสุดท้ายของชั้นที่ 2

$L = 12.5$, $U = 24.5$, $I = 12$

$F_L = 8$, $F_M = 20$, $f_M = 12$

$$\begin{aligned} \text{Med} &= 12.5 + \left(\tfrac{20-8}{12}\right)(12) \\ &= 12.5 + 12 \\ &= 24.5 \end{aligned}$$

หรือถ้าใช้อีกสูตร

$$\begin{aligned} \text{Med} &= 24.5 - \left(\tfrac{20-20}{12}\right)(12) \\ &= 24.5 - 0 \\ &= 24.5 \end{aligned}$$

เท่ากัน

สังเกตว่า ถ้าบังเอิญ Med ไปตกอยู่ "ตัวสุดท้ายของชั้น" แล้ว $U - \left(\frac{F_M - \frac{N}{2}}{f_M}\right)I$ จะเท่ากับ $U - 0 = U$ เสมอ

นักเรียนส่วนใหญ่จึงนิยมจำว่า "ตัวสุดท้ายของชั้น จะมีค่าเท่ากับขอบบนของชั้น" เสมอ

เช่น

คะแนนสอบ	จำนวนนักเรียน (f)	ความถี่สะสม (F)
1 - 12	8	8
13 - 24	12	20
25 - 36	10	30
37 - 48	7	37
49 - 60	3	40

$\Rightarrow$

ตัวที่ 8 มีค่า 12.5

ตัวที่ 20 มีค่า 24.5

ตัวที่ 30 มีค่า 36.5

ตัวที่ 37 มีค่า 48.5

ตัวที่ 40 มีค่า 60.5 เป็นต้น

<u>แบบฝึกหัด</u>

1. จงหามัธยฐานของข้อมูลต่อไปนี้

1.

คะแนน	ความถี่
1 - 10	6
11 - 20	12
21 - 30	10
31 - 40	2

2.

คะแนน	ความถี่
1 - 4	2
5 - 8	5
9 - 12	10
13 - 16	13

3.

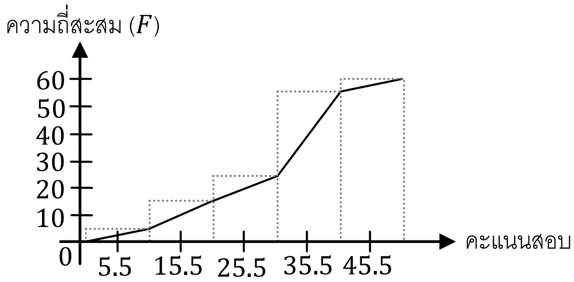
คะแนน	ความถี่สะสม
1 - 10	6
11 - 20	18
21 - 30	20
31 - 40	35
41 - 50	40

มัธยฐาน (จากกราฟ)

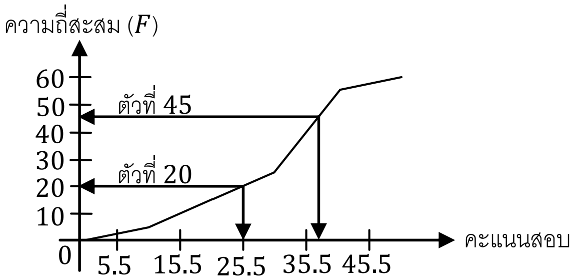
เราสามารถหา Med แบบคร่าวๆ จาก "โอจีฟ" ได้

โอจีฟ คือ กราฟที่ได้จากการนำ "ความถี่สะสม" มาเขียนกราฟ แล้วลากเส้นเชื่อมยอดแท่ง ดังรูป

คะแนนสอบ	จำนวนนักเรียน (คน)	ความถี่สะสม
1 - 10	3	3
11 - 20	12	15
21 - 30	10	25
31 - 40	29	54
41 - 50	6	60



ซึ่งจากหัวข้อที่เรียนไปก่อนหน้านี้ เราสามารถใช้โอจีฟ มาพล็อตหาค่าประมาณของข้อมูล ณ ตำแหน่งต่างๆได้

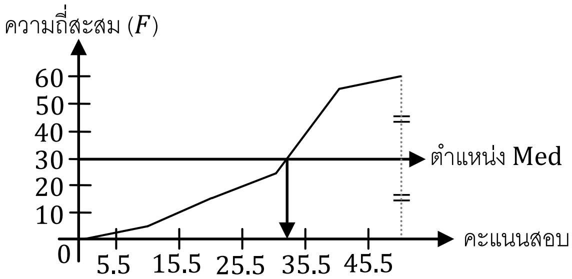


ในการหา Med จาก โอจีฟ เราจะหาตำแหน่งของ Med มาก่อน แล้วพล็อตหาค่าประมาณของตำแหน่งดังกล่าว

เนื่องจาก ความถี่สะสมช่องสุดท้าย จะมีค่าเท่ากับ N เสมอ

ดังนั้น ตำแหน่ง Med ที่เคยหาจากสูตร $\frac{N}{2}$ จะสามารถหาได้จาก $\frac{\text{ความสูงสุดท้ายของโอจีฟ}}{2}$ ด้วย

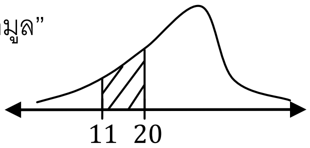
และเราสามารถใช้ตำแหน่งดังกล่าว พล็อตหา Med ได้ดังรูป



เช่น จากรูป จะได้ Med มีค่าประมาณ 32 กว่าๆ

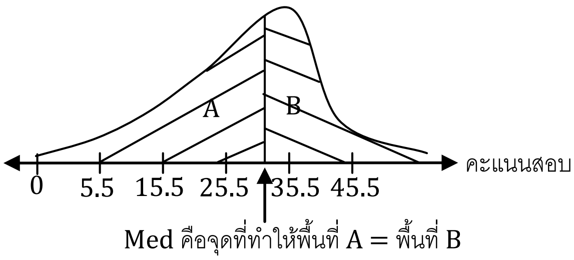
นอกจากนี้ เรายังสามารถหา Med แบบคร่าวๆ จาก “โค้งความถี่” ได้อีกด้วย

โค้งความถี่ จะมีสมบัติที่สำคัญอย่างหนึ่ง คือ “พื้นที่ใต้กราฟ” จะเป็นสัดส่วนกับ “จำนวนข้อมูล”
เช่น จำนวนนักเรียนที่ได้คะแนน 11 - 20 จะเป็นสัดส่วนกับพื้นที่ใต้โค้ง จาก 11 ถึง 20



เนื่องจาก Med เป็นจุดที่แบ่ง “จำนวนข้อมูล” ออกเป็น 2 ส่วน เท่าๆกัน
ดังนั้น Med จะเป็นจุดที่แบ่ง “พื้นที่ใต้โค้ง” เป็น 2 ส่วนเท่าๆกัน ด้วย

ดังนั้น ถ้าให้โค้งความถี่มา เราสามารถประมาณค่า Med โดยเล็งหาจุดที่แบ่งพื้นที่ใต้โค้งออกเป็น 2 ส่วนเท่ากัน ดังรูป



สมบัติของมัธยฐาน

สมบัติอย่างแรกของมัธยฐาน คือ มัน "ทน" กับข้อมูลที่มากผิดปกติ หรือน้อยผิดปกติ ได้
ปกติ เราจะไม่ค่อยชอบให้ข้อมูลที่ผิดปกติ มาฉุดภาพรวมของข้อมูลไปซักเท่าไหร่
เช่น $6, 7, 7, 8, 11, 12, 12, 2000$ → จะเห็นว่า 2000 เป็นข้อมูลที่มากผิดปกติ
ซึ่งการมีอยู่ของ 2000 นี้จะ "ฉุด $\bar{x}$" ได้สุดๆ

เช่น ถ้าไม่มี 2000: $6, 7, 7, 8, 11, 12, 12$ $\quad \bar{x} = \frac{6+7+7+8+11+12+12}{7} = 9$

ถ้า มี 2000: $6, 7, 7, 8, 11, 12, 12, 2000$ $\quad \bar{x} = \frac{6+7+7+8+11+12+12+2000}{8} = 257.88$

อย่างไรก็ตาม การมีอยู่ของ 2000 จะ ไม่มีผลกับ Med มากนัก

เช่น ถ้าไม่มี 2000: $6, 7, 7, 8, 11, 12, 12$ $\quad \text{Med} = 8$

ถ้า มี 2000: $6, 7, 7, 8, 11, 12, 12, 2000$ $\quad \text{Med} = \frac{8+11}{2} = 9.5$

ดังนั้น ถ้าในข้อมูล มีบางตัวที่มากหรือน้อยผิดปกติอยู่ เรามักนิยมใช้ Med เป็นค่ากลาง แทน $\bar{x}$

สมบัติอีกอย่างที่สำคัญ คือ $\sum|x_i - a|$ จะมีค่าน้อยที่สุด เมื่อ $a = \text{Med}$
ข้อนี้ จะคล้ายๆ กับของ $\bar{x}$ ซึ่งนักเรียนส่วนใหญ่ มักจะจำสับกัน

$\sum(x_i - a)^2$ จะมีค่าน้อยที่สุด เมื่อ $a = \bar{x}$

แต่ $\sum|x_i - a|$ จะมีค่าน้อยที่สุด เมื่อ $a = \text{Med}$

ตัวอย่างเช่น ข้อมูล $6, 7, 7, 8, 11, 12, 12$ มี $\text{Med} = 8$

$$\begin{aligned}\sum|x_i - 8| &= |6-8| + |7-8| + |7-8| + |8-8| + |11-8| + |12-8| + |12-8| \\ &= 2 + 1 + 1 + 0 + 3 + 4 + 4 \\ &= 15\end{aligned}$$

สมบัติของ Med จะรับประกันได้ว่า 15 จะเป็นค่าที่น้อยที่สุด ในบรรดา $\sum|x_i - a|$ ทั้งหมด
พูดง่ายๆก็คือ ถ้า a เป็นค่าอื่นที่ไม่ใช่ Med จะไม่มีวันได้ $\sum|x_i - a|$ น้อยกว่า 15
เช่น ถ้าแทน a ด้วย 7 จะได้

$$\begin{aligned}\sum|x_i - 7| &= |6-7| + |7-7| + |7-7| + |8-7| + |11-7| + |12-7| + |12-7| \\ &= 1 + 0 + 0 + 1 + 4 + 5 + 5 \\ &= 16 \quad \rightarrow \text{มากกว่า } 15\end{aligned}$$

หรือ ถ้าลองแทน a ด้วย $\bar{x}$ ดู (ข้อมูลชุดนี้ มี $\bar{x} = \frac{6+7+7+8+11+12+12}{7} = 9$) จะได้

$$\begin{aligned}\sum|x_i - 9| &= |6-9| + |7-9| + |7-9| + |8-9| + |11-9| + |12-9| + |12-9| \\ &= 3 + 2 + 2 + 1 + 2 + 3 + 3 \\ &= 16 \quad \rightarrow \text{ก็ยังแพ้ Med}\end{aligned}$$

ดังนั้น จะเห็นว่า $\sum|x_i - a|$ จะมีค่าน้อยที่สุด เมื่อ $a = \text{Med}$

<u>แบบฝึกหัด</u>

1. ให้ x เป็นจำนวนจริงใดๆ จงหาค่าต่ำสุดของ
    1. $|2-x|+|4-x|+|9-x|$
    2. $|x-2|+|x-4|+|x-9|+|x-10|$
    3. $|x+2|+|4-x|+|x+3|$
    4. $2|1-x|+3|x+4|+|x|$

ฐานนิยม

ค่ากลางตัวสุดท้ายที่ต้องเรียน คือ “ฐานนิยม”

ฐานนิยม (Mode) คือ ค่าที่ข้อมูลส่วนใหญ่นิยมเป็น

เราสามารถหาฐานนิยมโดยหาข้อมูลที่ “ซ้ำมากสุด”

- ถ้ามีข้อมูลที่ซ้ำมากสุดเท่ากัน สองตัว ก็ให้ตอบเป็นฐานนิยมไปทั้งสองตัว
- ถ้ามีข้อมูลที่ซ้ำมากสุดเท่ากัน เกินสองตัว นิยมตอบว่า “ไม่มีฐานนิยม”
- ถ้าข้อมูลไม่ซ้ำกันเลย ตอบว่า “ไม่มีฐานนิยม”

เช่น

$5, 6, 6, 7, 8, 8, 8, 10, 10, 11, 12, 12$
$\rightarrow$ Mode $= 8$

$6, 7, 7, 11, 12, 15, 17, 20, 20, 24, 28$
$\rightarrow$ Mode $= 7, 20$

$6, 6, 10, 11, 12, 12, 13, 15, 15, 16, 18$
$\rightarrow$ Mode $=$ ไม่มี

คะแนนสอบ	จำนวนนักเรียน (f)
11	6
12	12
13	15
14	18
15	2

$\rightarrow$ Mode $= 14$

อายุ	จำนวนนักเรียน (f)
14	25
15	53
16	31
17	53

$\rightarrow$ Mode $= 15, 17$

<u>แบบฝึกหัด</u>

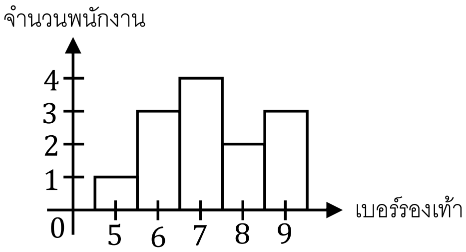
1. จงหาฐานนิยมของข้อมูลต่อไปนี้

   1. $3, 4, 8, 12, 19, 13, 16, 12$

   2.

0	|	2			
1	|	1	2	3	1
2	|	4	9	9	4
3	|	4			

   3.



   4.

คะแนน	ความถี่
7	2
8	8
9	12
10	6

   5.

คะแนน	ความถี่สะสม
23	2
24	7
25	12
26	13

2. ข้อมูลสองชุดเป็นดังนี้ ชุดที่ 1 : 1 3 3 6 8 9

ชุดที่ 2 : 2 3 4 5 5 5

ข้อใด <u>ผิด</u> [O-NET 58/31]

1. ค่าเฉลี่ยเลขคณิตของข้อมูลชุดที่ 1 มากกว่าค่าเฉลี่ยเลขคณิตของข้อมูลชุดที่ 2 อยู่ 0.5
2. ข้อมูลทั้งสองชุดมีมัธยฐานเท่ากัน
3. ฐานนิยมของข้อมูลสองชุดนี้ต่างกันอยู่ 2
4. ค่าเฉลี่ยเลขคณิตรวมของข้อมูลทั้งสองชุดเท่ากับ 4.5
5. ค่าเฉลี่ยเลขคณิตของข้อมูลชุดที่ 1 เท่ากับฐานนิยมของข้อมูลชุดที่ 2

3. แผนภาพต้นใบของข้อมูลชุดหนึ่งเป็นดังนี้

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

ข้อใดถูกต้องบ้าง [O-NET 54/20]

1. ข้อมูลชุดนี้ไม่มีฐานนิยม
2. มัธยฐานของข้อมูลชุดนี้เท่ากับ 40

4. ถ้าข้อมูลชุดหนึ่งประกอบด้วย 10, 12, 15, 13 และ 10 ข้อความในข้อใดต่อไปนี้เป็น<u>เท็จ</u> สำหรับข้อมูลชุดนี้ [O-NET 49/1-14]

1. มัธยฐาน เท่ากับ 12
2. ฐานนิยม น้อยกว่า 12
3. ฐานนิยม น้อยกว่า ค่าเฉลี่ยเลขคณิต
4. ค่าเฉลี่ยเลขคณิต มากกว่า 12

5. แผนภาพต้นใบของน้ำหนักในหน่วยกรัมของไข่ไก่ 10 ฟอง เป็นดังนี้

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

ข้อสรุปใดเป็น<u>เท็จ</u> [O-NET 53/29]

1. ฐานนิยมของน้ำหนักของไข่ไก่มีเพียงค่าเดียว
2. ค่าเฉลี่ยเลขคณิตและมัธยฐานของน้ำหนักของไข่ไก่มีค่าเท่ากัน
3. มีไข่ไก่ 5 ฟองที่มีน้ำหนักน้อยกว่า 70 กรัม
4. ไข่ไก่ที่มีน้ำหนักสูงกว่าฐานนิยม มีจำนวนมากกว่า ไข่ไก่ที่มีน้ำหนักเท่ากับฐานนิยม

6. ข้อมูลชุดหนึ่งประกอบด้วย

4, 9, 2, 7, 6, 5, 4, 6, 3, 4

ข้อใดต่อไปนี้<u>ถูกต้อง</u> [O-NET 52/27]

1. ค่าเฉลี่ยเลขคณิต < ฐานนิยม < มัธยฐาน
2. ฐานนิยม < มัธยฐาน < ค่าเฉลี่ยเลขคณิต
3. ฐานนิยม < ค่าเฉลี่ยเลขคณิต < มัธยฐาน
4. มัธยฐาน < ฐานนิยม < ค่าเฉลี่ยเลขคณิต

7. กำหนดแผนภาพ ต้น - ใบ ของข้อมูลชุดหนึ่ง ดังนี้

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

สำหรับข้อมูลชุดนี้ ข้อใดต่อไปนี้เป็นจริง [O-NET 51/36]

1. มัธยฐาน < ฐานนิยม < ค่าเฉลี่ยเลขคณิต
2. มัธยฐาน < ค่าเฉลี่ยเลขคณิต < ฐานนิยม
3. ค่าเฉลี่ยเลขคณิต < ฐานนิยม < มัธยฐาน
4. ค่าเฉลี่ยเลขคณิต < มัธยฐาน < ฐานนิยม

8. ถ้าน้ำหนัก (คิดเป็นกิโลกรัม) ของนักเรียน 2 กลุ่ม กลุ่มละ 6 คน เขียนเป็นแผนภาพ ต้น - ใบ ได้ดังนี้

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

ข้อสรุปในข้อใดต่อไปนี้ถูกต้อง [O-NET 49/1-30]

1. น้ำหนักเฉลี่ยของนักเรียนกลุ่มที่ 2 มากกว่าน้ำหนักเฉลี่ยของนักเรียนกลุ่มที่ 1
2. ฐานนิยมของน้ำหนักของนักเรียนกลุ่มที่ 2 มากว่าฐานนิยมของน้ำหนักของนักเรียนกลุ่มที่ 1
3. มัธยฐานของน้ำหนักของนักเรียนกลุ่มที่ 2 มากกว่ามัธยฐานของน้ำหนักของนักเรียนกลุ่มที่ 1
4. มัธยฐานของน้ำหนักของนักเรียนทั้งหมด มากกว่ามัธยฐานของน้ำหนักของนักเรียนกลุ่มที่ 1

9. จากแผนภาพ ต้น-ใบ ของข้อมูลแสดงน้ำหนัก (กิโลกรัม) ของนักเรียนกลุ่มหนึ่งเป็นดังนี้

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

เมื่อสุ่มเลือกนักเรียนมา 1 คน จากกลุ่มนี้ ความน่าจะเป็นที่จะได้นักเรียนที่มีน้ำหนักน้อยกว่าฐานนิยมของกลุ่ม มีค่าเท่ากับเท่าใด [O-NET 50/36]

10. ถ้าสุ่มตัวเลขหนึ่งจากข้อมูลชุดใดๆ ซึ่งประกอบด้วยตัวเลข 101 ตัว แล้ว ข้อใดต่อไปนี้ถูก [O-NET 51/39]

1. ความน่าจะเป็นที่ตัวเลขที่สุ่มได้มีค่าน้อยกว่าค่ามัธยฐาน $< \frac{1}{2}$
2. ความน่าจะเป็นที่ตัวเลขที่สุ่มได้มีค่าน้อยกว่าค่าเฉลี่ยเลขคณิต $< \frac{1}{2}$
3. ความน่าจะเป็นที่ตัวเลขที่สุ่มได้มีค่ามากกว่าค่ามัธยฐาน $> \frac{1}{2}$
4. ค่าเฉลี่ยเลขคณิต $<$ มัธยฐาน $<$ ฐานนิยม

ฐานนิยม (อันตรภาคชั้นเป็นช่วง)

ในกรณีที่ข้อมูลมาในรูปตารางแจกแจงความถี่ที่มีอันตรภาคชั้นเป็นช่วง ก็จะมีวิธีหา Mode ที่ยากขึ้นพอสมควร

คะแนนสอบ	จำนวนนักเรียน (f)
1 - 10	8
11 - 20	10
21 - 30	20
31 - 40	12

ในกรณีนี้ ถ้าจะหา Mode จะมีขั้นตอนดังนี้

1. หาอันตรภาคชั้นที่ Mode ตกอยู่ โดยดูจากชั้นที่ความถี่สูงสุด
2. คำนวณ “ผลต่างความถี่” ระหว่างชั้น Mode กับชั้นที่อยู่ติดกับ Mode ทั้งชั้นบนและชั้นล่าง

   d_1 = ความถี่ชั้น Mode − ความถี่ชั้นต่ำกว่า

   d_2 = ความถี่ชั้น Mode − ความถี่ชั้นสูงกว่า
3. คำนวณค่า Mode ด้วยสูตรไหนก็ได้ จาก 2 สูตรต่อไปนี้

ความกว้าง ชั้น Mode

$$\text{Mode} = L + \left(\frac{d_1}{d_1+d_2}\right) I$$

ขอบล่าง ชั้น Mode

ความกว้าง ชั้น Mode

$$\text{Mode} = U - \left(\frac{d_2}{d_1+d_2}\right) I$$

ขอบบน ชั้น Mode

เช่น

คะแนนสอบ	จำนวนนักเรียน (f)
1 - 10	8
11 - 20	10
21 - 30	20
31 - 40	12

ชั้นที่ 3 ความถี่มากสุด $= 20 \rightarrow$ Mode อยู่ชั้นที่ 3

$L = 20.5$, $U = 30.5$

$d_1 = 20 - 10 = 10$

$d_2 = 20 - 12 = 8$

$$\begin{aligned}\text{Mode} &= 20.5 + \left(\frac{10}{10+8}\right) 10 \\ &= 20.5 + 5.56 \\ &= 26.06\end{aligned}$$

หรือถ้าใช้อีกสูตร
$$\begin{aligned}\text{Mode} &= 30.5 - \left(\frac{8}{10+8}\right) 10 \\ &= 30.5 - 4.44 \\ &= 26.06\end{aligned}$$
เท่ากัน

ตัวอย่างอีกอัน เช่น

คะแนนสอบ	จำนวนนักเรียน (f)
1 - 8	16
9 - 16	12
17 - 24	11
25 - 32	9

ชั้นที่ 1 ความถี่มากสุด $= 16 \rightarrow$ Mode อยู่ชั้นที่ 1

$L = 0.5$, $U = 8.5$

$d_1 = 16 - 0 = 16$ (ถือว่าก่อนชั้น 1 มีความถี่ $= 0$)

$d_2 = 16 - 12 = 4$

$$\begin{aligned}\text{Mode} &= 0.5 + \left(\frac{16}{16+4}\right) 8 \\ &= 0.5 + 6.4 \\ &= 6.9\end{aligned}$$

หรือถ้าใช้อีกสูตร
$$\begin{aligned}\text{Mode} &= 8.5 - \left(\frac{4}{16+4}\right) 8 \\ &= 8.5 - 1.6 \\ &= 6.9\end{aligned}$$
เท่ากัน

ในกรณีที่ อันตรภาคชั้นกว้างไม่เท่ากัน เราจะไม่สามารถใช้ความถี่มาวัดความนิยมของชั้นได้อีกต่อไป
(ไม่งั้นชั้นที่แคบ จะเสียเปรียบชั้นที่กว้าง เพราะชั้นที่กว้างจะมีโอกาสที่มีข้อมูลอยู่หลายตัวกว่า)

ในกรณีนี้ เราจะใช้ ความหนาแน่นของชั้น $= \frac{\text{ความถี่ } (f)}{\text{ความกว้างชั้น } (I)}$ มาวัดความนิยม และใช้ในการคำนวณ d แทนความถี่ f

เช่น

คะแนนสอบ	จำนวนนักเรียน (f)	ความกว้างชั้น (I)	ความหนาแน่น $\frac{f}{I}$
1 – 8	8	8	1
9 – 10	10	2	5 → หนาแน่นสุด
11 – 20	20	10	2
21 – 35	12	15	0.8

ชั้นที่ 2 หนาแน่นสุด $= 5 \rightarrow$ Mode อยู่ชั้นที่ 2

$L = 8.5\,,\ I = 2$

$d_1 = 5 - 1 = 4$

$d_2 = 5 - 2 = 3$

$$\rightarrow \text{Mode} = 8.5 + \left(\frac{4}{4+3}\right)2$$
$$= 8.5 + 1.14$$
$$= 9.64$$

แบบฝึกหัด

1. จงหาฐานนิยมของข้อมูลต่อไปนี้

1.

คะแนน	ความถี่
1 - 10	3
11 - 20	9
21 - 30	12
31 - 40	11

2.

คะแนน	ความถี่
1 - 5	4
6 - 10	13
11 - 15	12
16 - 20	11

3.

คะแนน	ความถี่
1 - 4	2
5 - 8	3
9 - 12	6
13 - 16	11

4.

คะแนน	ความถี่สะสม
11 - 16	4
17 - 22	16
23 - 28	26
29 - 34	37
35 - 40	42

5.

คะแนน	ความถี่
1 - 10	5
11 - 30	10
31 - 40	6
41 - 60	6

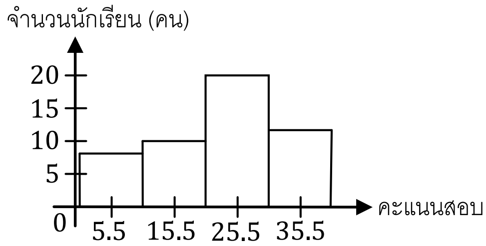
6.

คะแนน	ความถี่สะสม
1 - 10	5
11 - 15	15
16 - 19	21
20 - 29	27

ฐานนิยม (จากกราฟ)

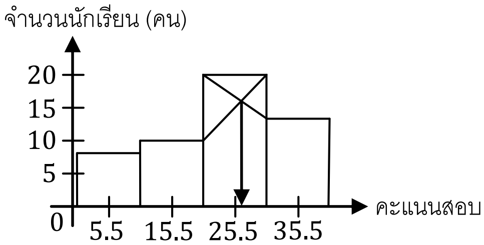
เราสามารถหา Mode จากฮิสโทแกรม ได้

คะแนนสอบ	จำนวนนักเรียน (f)
1 - 10	8
11 - 20	10
21 - 30	20
31 - 40	12



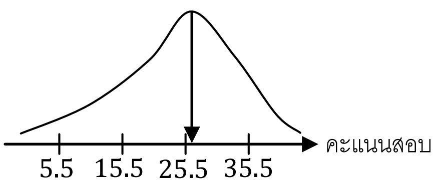
โดยมีขั้นตอน ดังนี้

1. หาแท่งที่สูงที่สุด
2. ลากเส้นไขว้จากยอดของแท่งที่สูงสุด ไปยังยอดของแท่งข้างๆ
3. จากจุดตัดที่ได้ ลากลงไปหาค่า Mode ที่แกนค่าข้อมูลได้เลย



เช่น เราจะหา Mode ได้จากตัวอย่างดังรูป
ซึ่งจะได้ Mode มีค่าประมาณ 26

นอกจากนี้ เรายังสามารถหา Mode จากโค้งความถี่ได้ด้วย
โดย Mode จะอยู่บริเวณที่โค้งความถี่ "โด่งที่สุด"



เช่น เราจะ Mode ได้จากตัวอย่างดังรูป
ซึ่งจะได้ Mode มีค่าประมาณ 26

สมบัติของฐานนิยม

ฐานนิยม เป็นค่ากลางเพียงตัวเดียว ที่สามารถใช้กับ “ข้อมูลเชิงคุณภาพ” ได้
ข้อมูลเชิงคุณภาพ คือ ข้อมูลที่แสดงคุณสมบัติ เช่น สีที่ชอบ เพศ กรุ๊ปเลือด

- ข้อมูลประเภทนี้ จะไม่สามารถ บวก ลบ คูณ หาร ได้ จึงหาค่าเฉลี่ยไม่ได้
- ข้อมูลประเภทนี้ จะไม่สามารถเทียบมากกว่าน้อยกว่า ได้ จึงหา Med ไม่ได้

เช่น ถ้าผลการสำรวจสีที่ชอบของนักเรียน 7 คน พบว่าชอบ สีฟ้า , สีเขียว , สีเขียว , สีชมพู , สีดำ , สีฟ้า , สีเขียว
ข้อมูลประเภทนี้ หมดสิทธิ์หา $\bar{x}$ หรือ Med เพราะ บวก ลบ คูณ หาร หรือ เรียงลำดับ ไม่ได้
แต่จะหา Mode ได้ $=$ สีเขียว

<u>แบบฝึกหัด</u>

1. การเลือกใช้ค่ากลางของข้อมูลควรพิจารณาสิ่งต่อไปนี้<u>ยกเว้น</u>ข้อใด [O-NET 52/29]
   1. ลักษณะของข้อมูล
   2. วิธีจัดเรียงลำดับข้อมูล
   3. จุดประสงค์ของการนำไปใช้
   4. ข้อดีและข้อเสียของค่ากลางแต่ละชนิด

2. ข้อใดถูก [O-NET 58/29]
   1. ข้อมูลที่จะวัดค่ากลางได้ต้องเป็นข้อมูลเชิงปริมาณเท่านั้น
   2. กรณีที่ข้อมูลมีจำนวนน้อยควรใช้ฐานนิยมเป็นค่ากลางเพราะสามารถนับความถี่ของข้อมูลได้สะดวก
   3. ค่าเฉลี่ยเลขคณิตเป็นค่ากลางที่ไม่เหมาะสมกับข้อมูลที่มีบางค่าต่ำกว่าข้อมูลอื่นๆมาก
   4. เนื่องจากมัธยฐานคือค่าของข้อมูลที่อยู่กึ่งกลางของข้อมูลทั้งชุด ดังนั้น มัธยฐานจึงใช้เฉพาะกรณีที่ข้อมูลมีจำนวนข้อมูลเป็นจำนวนคี่เท่านั้น
   5. ค่ากลางของข้อมูลที่แจกแจงความถี่แล้วมีความถูกต้องแน่นอนมากกว่าค่ากลางของข้อมูลชุดเดียวกันที่ยังไม่ได้แจกแจงความถี่

3. ค่ากลางของข้อมูลในข้อใดมีความเหมาะสมที่จะใช้เป็นตัวแทนของข้อมูลของกลุ่ม [O-NET 56/29]
   1. ค่าเฉลี่ยเลขคณิตของน้ำหนักตัวของชาวจังหวัดเชียงใหม่
   2. ค่าเฉลี่ยเลขคณิตของจำนวนหน้าของหนังสือที่คนไทยแต่ละคนอ่านในปี พ.ศ. 2554
   3. มัธยฐานของจำนวนเงินที่แต่ละคนใช้จ่ายต่อเดือนของคนไทย
   4. ฐานนิยมของความสูงของนักเรียนห้องหนึ่ง
   5. ค่าเฉลี่ยของฐานนิยมกับมัธยฐานของคะแนนสอบของนักเรียนทั้งโรงเรียน

4. ค่ากลางของข้อมูลในข้อใดมีความเหมาะสมที่จะใช้เป็นตัวแทนของข้อมูลของกลุ่ม [O-NET 57/29]
   1. ค่าเฉลี่ยเลขคณิตของขนาดรองเท้าของนักเรียนห้องหนึ่ง
   2. ค่าเฉลี่ยเลขคณิตของจำนวนผู้โดยสารรถไฟฟ้าใต้ดินต่อวันในเดือน มกราคม พ.ศ. 2557
   3. มัธยฐานของน้ำหนักตัวของคนไทยในปี พ.ศ. 2556
   4. ฐานนิยมของความสูงของนักกีฬาไทยได้ที่ได้รับเหรียญทองจากการแข่งขันกีฬาโอลิมปิก
   5. ค่ากึ่งกลางระหว่างมัธยฐานกับค่าเฉลี่ยเลขคณิตของคะแนนสอบวิชาคณิตศาสตร์ของนักเรียนชั้น ม.6 ของโรงเรียนแห่งหนึ่ง

5. ข้อมูลต่อไปนี้แสดงน้ำหนักในหน่วยกิโลกรัม ของนักเรียนกลุ่มหนึ่ง

   $41, 88, 46, 42, 43, 49, 44, 45, 43, 95, 47, 48$

   ค่ากลางในข้อใดเป็นค่าที่เหมาะสมที่จะเป็นตัวแทนของข้อมูลชุดนี้ [O-NET 53/31]
   1. มัธยฐาน
   2. ฐานนิยม
   3. ค่าเฉลี่ยเลขคณิต
   4. ค่าเฉลี่ยของค่าสูงสุดและค่าต่ำสุด

6. ยอดขายต่อเดือน (หน่วย : หมื่นบาท) ของบริษัทแห่งหนึ่งในระยะเวลา 10 เดือน เป็นดังนี้

   $154 \quad 151 \quad 148 \quad 405 \quad 158 \quad 157 \quad 158 \quad 148 \quad 148 \quad 153$

   ข้อใดถูก [O-NET 59/29]
   1. ค่าเฉลี่ยเลขคณิต $(\bar{x})$ เป็นค่ากลางที่เหมาะสมที่สุดสำหรับเป็นตัวแทนของข้อมูลนี้ และ $\bar{x} = 178$
   2. ฐานนิยม เป็นค่ากลางที่เหมาะสมที่สุดสำหรับเป็นตัวแทนของข้อมูลนี้ และ ฐานนิยม $= 148$
   3. ฐานนิยม เป็นค่ากลางที่เหมาะสมที่สุดสำหรับเป็นตัวแทนของข้อมูลนี้ และ ฐานนิยม $= 158$
   4. มัธยฐาน เป็นค่ากลางที่เหมาะสมที่สุดสำหรับเป็นตัวแทนของข้อมูลนี้ และ มัธยฐาน $= 157.5$
   5. มัธยฐาน เป็นค่ากลางที่เหมาะสมที่สุดสำหรับเป็นตัวแทนของข้อมูลนี้ และ มัธยฐาน $= 153.5$

ค่าเฉลี่ย - มัธยฐาน - ฐานนิยม

หัวข้อนี้ จะพูดถึง ความสัมพันธ์ระหว่างค่ากลางทั้ง 3 ตัว

จากการสังเกตจากข้อมูลที่มีอยู่จริงๆ เราพบว่า ค่าเฉลี่ย มัธยฐาน และ ฐานนิยม จะมีความสัมพันธ์กันอยู่นิดๆ

ความสัมพันธ์ดังกล่าว คือ

$$\boxed{\bar{x} - \text{Mode} = 3(\bar{x} - \text{Med})}$$

ก่อนอื่น ต้องเข้าใจก่อนว่า สูตรนี้ ได้จากการสังเกตจากข้อมูลจริง

อาจจะไม่ได้ผลถูกต้อง 100% แต่ก็สามารถใช้ประมาณของค่ากลาง เมื่อรู้ค่ากลางอีก 2 ชนิดได้

<u>ตัวอย่าง</u> ข้อมูลชุดหนึ่ง มี $\bar{x} = 20$ และ $\text{Med} = 18$ จงหาค่าประมาณของฐานนิยม

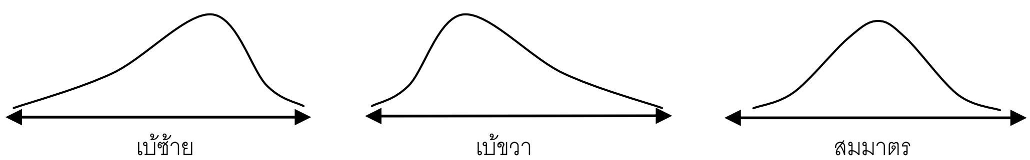
<u>วิธีทำ</u>

$$\begin{aligned} \bar{x} - \text{Mode} &= 3(\bar{x} - \text{Med}) \\ 20 - \text{Mode} &= 3(20 - 18) \\ 20 - \text{Mode} &= 6 \\ \text{Mode} &= 14 \end{aligned}$$

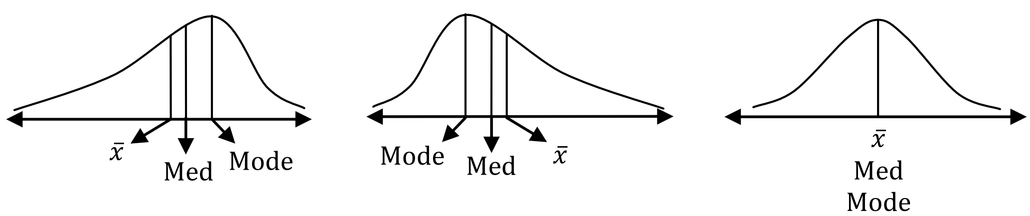
#

นอกจากนี้ ลำดับมากน้อยของค่ากลางทั้ง 3 ตัว ยังสามารถบอก ลักษณะ “การเบ้” ได้ด้วย

ตอนที่เรียนเรื่องโค้งความถี่ เราแบ่งประเภทโค้งความถี่ได้ 3 แบบ ตามลักษณะการเบ้ ดังนี้



ซึ่งในการเบ้แต่ละแบบ ค่ากลางทั้ง 3 ตัว จะเรียงลำดับมากน้อยไม่เหมือนกัน ดังนี้



สิ่งที่ต้องจำ คือ ในโค้งเบ้ซ้าย $\bar{x} < \text{Med} < \text{Mode}$

ในโค้งเบ้ขวา $\text{Mode} < \text{Med} < \bar{x}$

ในโค้งสมมาตร $\bar{x} = \text{Med} = \text{Mode}$

หมายเหตุ : ข้อมูลที่สำรวจจากธรรมชาติจำนวนมากๆ จะมีลักษณะการแจกแจงเฉพาะตัว ที่เรียกว่า “การแจกแจงปกติ”
ซึ่งข้อมูลจากธรรมชาตินี้ จะมีลักษณะเป็นโค้งสมมาตรเสมอ

<u>ตัวอย่าง</u> ข้อมูลชุดหนึ่ง มีมัธยฐานเท่ากับ 24 และฐานนิยมเท่ากับ 18 จงหาค่าประมาณของ ค่าเฉลี่ย พร้อมทั้งระบุว่า ข้อมูลชุดนี้ มีลักษณะการเบ้แบบใด

<u>วิธีทำ</u>

$$\begin{aligned} \bar{x} - \text{Mode} &= 3(\bar{x} - \text{Med}) \\ \bar{x} - 18 &= 3(\bar{x} - 24) \\ \bar{x} - 18 &= 3\bar{x} - 72 \\ 54 &= 2\bar{x} \\ \bar{x} &= 27 \end{aligned}$$

เนื่องจาก $18 < 24 < 27$ ดังนั้น $\text{Mode} < \text{Med} < \bar{x}$

ดังนั้น ข้อมูลชุดนี้มีลักษณะการกระจายแบบเบ้ขวา #

<u>แบบฝึกหัด</u>

1. จงหาลักษณะการเบ้ของข้อมูลต่อไปนี้

1. 2 , 3 , 4 , 6 , 10 , 10 , 10 , 11

2.

0	2	4		
1	3	3	4	5
2	1	2		
3	1			

2. จงหาค่าประมาณของค่ากลางที่เหลือ พร้อมทั้งระบุลักษณะการเบ้

1. $\bar{x} = 5$, $\text{Mode} = 11$

2. $\bar{x} = 11$, $\text{Med} = 9$

3. $\text{Mode} = 12$, $\text{Med} = 14$

3 คะแนนสอบของนักเรียนห้องหนึ่ง มีการแจกแจงแบบปกติ ถ้าห้องนี้มีนักเรียน 40 คน และฐานนิยมเท่ากับ 60 คะแนน แล้ว จงหาผลรวมคะแนนของนักเรียนทั้งห้อง

4. ข้อมูลชุดหนึ่งมีค่าเฉลี่ยเลขคณิตเท่ากับ 20 มัธยฐานเท่ากับ 25 และฐานนิยมเท่ากับ 30 ข้อสรุปใดต่อไปนี้ถูกต้อง [O-NET 52/30]
   1. ลักษณะการกระจายของข้อมูลเป็นการกระจายที่เบ้ทางซ้าย
   2. ลักษณะการกระจายของข้อมูลเป็นการกระจายที่เบ้ทางขวา
   3. ลักษณะการกระจายของข้อมูลเป็นการกระจายแบบสมมาตร
   4. ไม่สามารถสรุปลักษณะการกระจายของข้อมูลได้

การวัดตำแหน่งข้อมูล

ในหัวข้อก่อนหน้า เราได้เรียนเกี่ยวกับ มัธยฐาน ซึ่งเป็นวิธีวัดตำแหน่งข้อมูลอย่างหนึ่ง
มัธยฐาน จะหมายถึง ข้อมูลที่มี “ตำแหน่งตรงกลาง” นั่นเอง

$$10 \quad 11 \quad 15 \quad 17 \quad 18 \quad 21 \quad 26 \quad 27 \quad 32 \quad \overset{\overset{\text{Med}}{\downarrow}}{\underline{33}} \quad 35 \quad 39 \quad 40 \quad 43 \quad 44 \quad 45 \quad 46 \quad 49 \quad 51$$

ในเรื่องนี้ เราจะได้เรียนข้อมูลในตำแหน่งอื่นๆ ซึ่งเราวัดได้โดยใช้ ควอไทล์ เดไซล์ และ เปอร์เซ็นไทล์

<u>ควอไทล์</u>

ถ้าเราแบ่งข้อมูลออกเป็น 4 ส่วนเท่าๆกัน จะมีจุดแบ่งอยู่ 3 จุด
เราจะเรียกข้อมูลตรงจุดแบ่งทั้ง 3 จุด ว่า Q_1 Q_2 และ Q_3 ตามลำดับ ดังนี้

$$10 \quad 11 \quad 15 \quad 17 \quad \overset{\overset{\mathrm{Q}_1}{\downarrow}}{\underline{18}} \quad 21 \quad 26 \quad 27 \quad 32 \quad \overset{\overset{\mathrm{Q}_2}{\downarrow}}{\underline{33}} \quad 35 \quad 39 \quad 40 \quad 43 \quad \overset{\overset{\mathrm{Q}_3}{\downarrow}}{\underline{44}} \quad 45 \quad 46 \quad 49 \quad 51$$

หมายเหตุ : จะเห็นว่า $\mathrm{Q}_2 = \mathrm{Med}$ เสมอ

ตำแหน่งของ Q_r จะหาได้จากสูตร $\frac{r}{4} \cdot (N+1)$
เช่น จากตัวอย่างด้านบน มีข้อมูล 19 ตัว ($N = 19$)
จะได้ Q_1 อยู่ตัวที่ $\frac{1}{4} \cdot (19+1) = 5$ $\quad$ Q_2 อยู่ตัวที่ $\frac{2}{4} \cdot (19+1) = 10$
Q_3 อยู่ตัวที่ $\frac{3}{4} \cdot (19+1) = 15$

<u>เดไซล์</u>

ถ้าเราแบ่งข้อมูลออกเป็น 10 ส่วนเท่าๆกัน จะมีจุดแบ่งอยู่ 9 จุด
เราจะเรียกข้อมูลตรงจุดแบ่งทั้ง 9 จุด ว่า D_1 D_2 D_3 ... D_9 ตามลำดับ ดังนี้

$$10 \quad \overset{\overset{\mathrm{D}_1}{\downarrow}}{\underline{11}} \quad 15 \quad \overset{\overset{\mathrm{D}_2}{\downarrow}}{\underline{17}} \quad 18 \quad \overset{\overset{\mathrm{D}_3}{\downarrow}}{\underline{21}} \quad 26 \quad \overset{\overset{\mathrm{D}_4}{\downarrow}}{\underline{27}} \quad 32 \quad \overset{\overset{\mathrm{D}_5}{\downarrow}}{\underline{33}} \quad 35 \quad \overset{\overset{\mathrm{D}_6}{\downarrow}}{\underline{39}} \quad 40 \quad \overset{\overset{\mathrm{D}_7}{\downarrow}}{\underline{43}} \quad 44 \quad \overset{\overset{\mathrm{D}_8}{\downarrow}}{\underline{45}} \quad 46 \quad \overset{\overset{\mathrm{D}_9}{\downarrow}}{\underline{49}} \quad 51$$

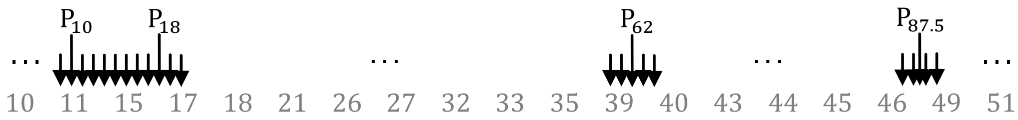
หมายเหตุ : จะเห็นว่า $\mathrm{D}_5 = \mathrm{Q}_2 = \mathrm{Med}$ เสมอ

ตำแหน่งของ D_r จะหาได้จากสูตร $\frac{r}{10} \cdot (N+1)$
เช่น จากตัวอย่างด้านบน มีข้อมูล 19 ตัว ($N = 19$)
จะได้ D_1 อยู่ตัวที่ $\frac{1}{10} \cdot (19+1) = 2$ $\quad$ D_2 อยู่ตัวที่ $\frac{2}{10} \cdot (19+1) = 4$
D_5 อยู่ตัวที่ $\frac{5}{10} \cdot (19+1) = 10$ $\quad$ D_9 อยู่ตัวที่ $\frac{9}{10} \cdot (19+1) = 18$ เป็นต้น

<u>เปอร์เซ็นไทล์</u>

ถ้าเราแบ่งข้อมูลออกเป็น 100 ส่วนเท่าๆกัน จะมีจุดแบ่งอยู่ 99 จุด

เราจะเรียกข้อมูลตรงจุดแบ่งทั้ง 99 จุด ว่า P_1 P_2 P_3 ... P_{99} ตามลำดับ



หมายเหตุ : จะเห็นว่า $\mathrm{P}_{10} = \mathrm{D}_1$ $\mathrm{P}_{20} = \mathrm{D}_2$ $\mathrm{P}_{30} = \mathrm{D}_3$ $\mathrm{P}_{40} = \mathrm{D}_4$ $\mathrm{P}_{50} = \mathrm{D}_5$

$\mathrm{P}_{60} = \mathrm{D}_6$ $\mathrm{P}_{70} = \mathrm{D}_7$ $\mathrm{P}_{80} = \mathrm{D}_8$ $\mathrm{P}_{90} = \mathrm{D}_9$

และ $\mathrm{P}_{25} = \mathrm{Q}_1$ $\mathrm{P}_{50} = \mathrm{Q}_2$ $\mathrm{P}_{75} = \mathrm{Q}_3$

และ $\mathrm{P}_{50} = \mathrm{Med}$

ตำแหน่งของ P_r จะหาได้จากสูตร $\frac{r}{100} \cdot (N+1)$

เช่น จากตัวอย่างด้านบน มีข้อมูล 19 ตัว ($N = 19$)

จะได้ P_{10} อยู่ตัวที่ $\frac{10}{100} \cdot (19+1) = 2$ $\qquad$ P_{18} อยู่ตัวที่ $\frac{18}{100} \cdot (19+1) = 3.6$

P_{62} อยู่ตัวที่ $\frac{62}{100} \cdot (19+1) = 12.4$ $\qquad$ $\mathrm{P}_{87.5}$ อยู่ตัวที่ $\frac{87.5}{100} \cdot (19+1) = 17.5$ เป็นต้น

ในการหาค่าของ ควอไทล์ เดไซล์ และ เปอร์เซ็นไทล์ จะมีขั้นตอน คล้ายๆตอนที่เราหา Med ดังนี้

1. หา "ตำแหน่ง"

$\boxed{\mathrm{Q}_r \to \frac{r}{4} \cdot (N+1)}$ $\quad$ $\boxed{\mathrm{D}_r \to \frac{r}{10} \cdot (N+1)}$ $\quad$ $\boxed{\mathrm{P}_r \to \frac{r}{100} \cdot (N+1)}$

2. เอาตำแหน่งที่ได้ ไปหา "ค่า"

   โดยต้องเรียงข้อมูลจากน้อยไปมาก นับจนถึงตำแหน่งที่ต้องการ แล้วตอบค่าข้อมูล ณ ตำแหน่งนั้น

   ถ้าตำแหน่งไม่เป็นจำนวนเต็ม ให้ประมาณเอาจากสองตัวที่คร่อมตำแหน่งนั้นอยู่

   เช่น ถ้าต้องการหาค่าข้อมูล ตัวที่ 5.62 จะหาได้ดังนี้

$$\text{ข้อมูลตัวที่ } 5.\underline{62} = \text{ตัวที่ } 5 + 0.62 \times (\text{ตัวที่ } 6 - \text{ตัวที่ } 5)$$

เราต้องหัดใช้สูตรนี้ให้คล่อง เช่น ตัวที่ 10.3 = ตัวที่ 10 + 0.3 × (ตัวที่ 11 − ตัวที่ 10)

ตัวที่ 8.03 = ตัวที่ 8 + 0.03 × (ตัวที่ 9 − ตัวที่ 8)

ตัวที่ 3.45 = ตัวที่ 3 + 0.45 × (ตัวที่ 4 − ตัวที่ 3) เป็นต้น

<u>ตัวอย่าง</u> จงหา Q_3 , D_2 และ P_{83} ของข้อมูล 15 , 18 , 18 , 19 , 20 , 24 , 24

<u>วิธีทำ</u> มีข้อมูล 7 ตัว → Q_3 จะอยู่ตัวที่ $\frac{3}{4}\cdot(7+1) = 6$

นับดู จะได้ตัวที่ 6 มีค่า 24 ดังนั้น $Q_3 = 24$

→ D_2 อยู่ตัวที่ $\frac{2}{10}\cdot(7+1) = 1.6$

ตัวที่ 1.6 = ตัวที่ 1 + 0.6 × (ตัวที่ 2 − ตัวที่ 1)

$= 15 + 0.6\times(18 - 15) = 16.8$

ดังนั้น $D_2 = 16.8$

→ P_{83} อยู่ตัวที่ $\frac{83}{100}\cdot(7+1) = 6.64$

ตัวที่ 6.64 จะอยู่ระหว่างตัวที่ 6 กับตัวที่ 7 แต่บังเอิญตัวที่ 6 กับตัวที่ 7 เท่ากัน = 24

ดังนั้น ไม่ต้องคิดเลข ตอบได้เลยว่า $P_{83} = 24$ #

สุดท้าย ถ้าเราเจอประโยคพวกนี้ ต้องรู้ว่าเป็นเรื่อง ควอไทล์ เดไซล์ เปอร์เซ็นไทล์

"มีข้อมูลประมาณ 3 ใน 4 ได้คะแนนน้อยกว่า สมชาย" → สมชาย $= Q_3$

"มีข้อมูลประมาณครึ่งหนึ่ง ได้คะแนนน้อยกว่า สมหญิง" → สมหญิง $=$ Med

"มีข้อมูลประมาณ 7 ใน 10 ได้คะแนนน้อยกว่า สมหวัง" → สมหวัง $= D_7$

"มีข้อมูลประมาณ 60% ได้คะแนน<u>มากกว่า</u> สมปอง" → สมปอง $= P_{40}$

"สมศรี ได้คะแนนสูงที่สุด ในกลุ่ม 20% ที่ได้คะแนนต่ำสุด" → สมศรี $= P_{20}$

"สมหมาย ได้คะแนนต่ำที่สุด ในกลุ่ม 30% ที่ได้คะแนนสูงสุด" → สมหมาย $= P_{70}$

"โอกาสที่จะสุ่มได้คนที่คะแนนน้อยกว่าสมบัติ เท่ากับ 3 ใน 10" → สมบัติ $= D_3$

สิ่งที่ต้องรู้ เวลาใช้คำศัพท์พวกนี้ คือ เราจะไม่จุกจิกเรื่อง "เท่ากับ"

นั่นคือ เราจะถือว่าประโยคต่างๆต่อไปนี้ อนุโลมให้มีความหมายเหมือนกันได้

"มีข้อมูลประมาณ 3 ใน 4 ได้คะแนนน้อยกว่า สมชาย"

"มีข้อมูลประมาณ 3 ใน 4 ได้คะแนนน้อยกว่า<u>หรือเท่ากับ</u> สมชาย"

"มีข้อมูลประมาณ 1 ใน 4 ได้คะแนนมากกว่า สมชาย"

"มีข้อมูลประมาณ 1 ใน 4 ได้คะแนนมากกว่า<u>หรือเท่ากับ</u> สมชาย"

อย่างไรก็ตาม การไม่จุกจิกแบบนี้ สร้างปัญหาให้เราได้พอสมควร

เพราะบางที เราต้องเดาเอาเอง ว่าโจทย์หมายถึงแบบไหน ว่าโจทย์จะนับสมชายเป็นหนึ่งในข้อมูลหรือไม่

ส่วนใหญ่ มักจะต้องเลือกแบบที่คิดแล้วมีตัวเลือกให้กา และ / หรือ ได้คำตอบเป็นเลขลงตัว

<u>ตัวอย่าง</u> จากข้อมูลคะแนนสอบต่อไปนี้

20	22	25	25	27	29	31	33	33	34
38	39	39	41	45	48	49	51	54	55
57	59	59							

จงหาว่า นักเรียนที่ได้คะแนนมากที่สุดในกลุ่ม 10% ต่ำสุด ได้คะแนนน้อยกว่านักเรียนที่ได้คะแนนน้อยที่สุดในกลุ่ม 10% สูงสุด อยู่กี่คะแนน

<u>วิธีทำ</u> คะแนนมากที่สุดในกลุ่ม 10% ต่ำสุด $\rightarrow P_{10}$

คะแนนน้อยที่สุดในกลุ่ม 10% สูงสุด $\rightarrow P_{90}$ ข้อนี้ โจทย์ให้หา $P_{90} - P_{10}$ นั่นเอง

นับดู จะมีข้อมูลทั้งหมด 23 ตัว $\rightarrow$ P_{10} จะอยู่ตัวที่ $\frac{10}{100} \cdot (23+1) = 2.4$

ตัวที่ 2.4 = ตัวที่ 2 + 0.4 × (ตัวที่ 3 − ตัวที่ 2)

$= 22 + 0.4 \times (25 - 22) = 23.2$

$\rightarrow$ P_{90} จะอยู่ตัวที่ $\frac{90}{100} \cdot (23+1) = 21.6$

ตัวที่ 21.6 = ตัวที่ 21 + 0.6 × (ตัวที่ 22 − ตัวที่ 21)

$= 57 + 0.6 \times (59 - 57) = 58.2$

ดังนั้น P_{10} น้อยกว่า P_{90} อยู่ $58.2 - 23.2 = 35$ #

<u>แบบฝึกหัด</u>

1. จงหาค่าในแต่ละข้อต่อไปนี้ เมื่อกำหนดข้อมูลให้ดังนี้

10	12	15	20	25	26	28	29
33	34	36	38	40	42	42	

1. Med
2. Q_3
3. D_3
4. P_{15}

2. ข้อมูลชุดหนึ่งมี 10 จำนวนประกอบด้วยจำนวนต่อไปนี้

4, 8, 8, 9, 14, 15, 18, 18, 22, 25

ควอไทล์ที่สามของข้อมูลชุดนี้มีค่าเท่ากับเท่าใด [O-NET 49/2-5]

3. คะแนนสอบวิชาคณิตศาสตร์ของนักเรียนห้องหนึ่งแสดงด้วยแผนภาพต้น - ใบได้ดังนี้

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

เปอร์เซ็นต์ไทล์ที่ 50 ของคะแนนสอบนี้เท่ากับคะแนนเท่าใด [O-NET 54/40]

4. จากแผนภาพต้น–ใบของข้อมูลชุดหนึ่งเป็นดังนี้

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

เปอร์เซ็นไทล์ที่ 86 ของข้อมูลชุดนี้เท่ากับเท่าใด [O-NET 58/40]

5. พิจารณาข้อมูลต่อไปนี้

10, 5, 6, 9, 12, 15, 8, 18

ค่าของ P_{80} ใกล้เคียงกับข้อใดต่อไปนี้<u>มากที่สุด</u> [O-NET 52/31]

1. 15.1 2. 15.4 3. 15.7 4. 16.0

6. บริษัทขนส่งพัสดุแห่งหนึ่งได้บันทึกระยะทาง (หน่วย : กิโลเมตร) ในการส่งของในแต่ละวัน เป็นเวลา 30 วัน เมื่อเรียงลำดับข้อมูลจากน้อยไปมาก ดังนี้

33	37	43	44	44	55	58	65	65	66
71	74	75	75	78	81	81	81	82	84
86	86	87	89	89	92	92	93	93	95

แล้ว เปอร์เซนไทล์ที่ 33 ของข้อมูลชุดนี้ เท่ากับเท่าใด [O-NET 59/28]

7. ผลการสอบวิชาคณิตศาสตร์ของนักเรียนห้องหนึ่ง (เรียงจากน้อยไปมาก) เป็นดังนี้

29, 35, 36, 40, 41, 43, 47, 50, 56, 59,
60, 61, 63, 65, 72, 72, 74, 75, 75, 78,
78, 78, 80, 80, 81, 82, 84, 87, 88, 89,
90, 90, 91, 91, 91, 92, 95, 95, 95, 97

เปอร์เซ็นไทล์ที่ 70 ของคะแนนสอบนี้เท่ากับเท่าใด [O-NET 57/31]

8. ข้อมูลชุดหนึ่งเรียงจากน้อยไปมากดังนี้ 5 10 12 20 x 26 30 42 47 y
ถ้าข้อมูลชุดนี้มีพิสัยเท่ากับ 45 และค่าเฉลี่ยเลขคณิตเท่ากับ 26.4 แล้ว ควอไทล์ที่สองของข้อมูลชุดนี้เท่ากับเท่าใด [O-NET 57/32]

9. คะแนนของผู้เข้าสอบ 15 คน เป็นดังนี้

45, 54, 59, 60, 62, 64, 65, 68, 70, 72, 73, 75, 76, 80, 81

ถ้าเกณฑ์ในการสอบผ่าน คือ ต้องได้คะแนนไม่ต่ำกว่าเปอร์เซ็นไทล์ที่ 60 แล้ว
ข้อใดต่อไปนี้เป็นคะแนนต่ำสุดของผู้ที่สอบผ่าน [O-NET 51/18]

1. 68 คะแนน
2. 70 คะแนน
3. 72 คะแนน
4. 73 คะแนน

10. ข้อมูลชุดหนึ่งประกอบด้วย 19 จำนวน ต่อไปนี้

6	8	9	12	12	15	15	16	18	19
20	20	21	22	23	24	25	30	30	

ควอไทล์ที่ 3 มีค่าต่างจากเปอร์เซ็นต์ไทล์ที่ 45 เท่ากับเท่าใด [O-NET 50/15]

11. ข้อมูลชุดหนึ่งมี 5 จำนวน ถ้าควอไทล์ที่หนึ่ง ควอไทล์ที่สอง และควอไทล์ที่สามเท่ากับ 18, 25, และ 28 ตามลำดับ แล้ว ค่าเฉลี่ยเลขคณิตของข้อมูลชุดนี้เท่ากับเท่าใด [O-NET 50/38]

12. ข้อมูลชุดหนึ่ง ถ้าเรียงจากน้อยไปมากแล้ว ได้เป็นลำดับเลขคณิตต่อไปนี้

2, 5, 8, ..., 92

ควอไทล์ที่ 3 ของข้อมูลชุดนี้มีค่าเท่ากับเท่าใด [O-NET 51/34]

13. คะแนนสอบวิชาวิทยาศาสตร์ของนักเรียนห้องหนึ่งจำนวน 119 คน เป็นดังนี้

คะแนนที่ได้	จำนวนนักเรียน (คน)
52	13
55	12
57	17
60	9
62	10
65	6
70	14
75	14
78	7
80	10
82	7

คะแนนที่เปอร์เซ็นไทล์ที่ 56 เท่ากับเท่าใด [O-NET 56/39]

14. นักเรียนกลุ่มหนึ่งจำนวน 80 คน ซึ่งมี ลำเจียก ลำดวน และลำพู รวมอยู่ด้วย ปรากฏผลการสอบดังนี้

    ลำดวนได้คะแนนตรงกับควอไทล์ที่สาม

    ลำพูได้คะแนนตรงกับเปอร์เซ็นไทล์ที่ 50

    ลำเจียกได้คะแนนเป็นลำดับที่ 30 เมื่อเรียงคะแนนจากมากไปหาน้อย

    ข้อใดต่อไปนี้เป็นการเรียงรายชื่อของผู้ที่ได้คะแนนน้อยไปหาผู้ที่ได้คะแนนมาก [O-NET 51/19]

    1. ลำพู ลำเจียก ลำดวน
    2. ลำพู ลำดวน ลำเจียก
    3. ลำเจียก ลำพู ลำดวน
    4. ลำเจียก ลำดวน ลำพู

15. จากการตรวจสอบลำดับที่ของคะแนนสอบของนาย ก และนาย ข ใน วิชาคณิตศาสตร์ที่มีผู้เข้าสอบ 400 คน ปรากฏว่านาย ก สอบได้คะแนนอยู่ในตำแหน่งควอไทล์ที่ 3 และนาย ข สอบได้คะแนนอยู่ในตำแหน่งเปอร์เซ็นไทล์ที่ 60 จำนวนนักเรียนที่สอบได้คะแนนระหว่างคะแนนของนาย ก และนาย ข มีประมาณกี่คน [O-NET 53/33]

16. เมื่อพิจารณาผลการสอบวิชาคณิตศาสตร์ของนักเรียน 39 คน พบว่า เปอร์เซ็นไทล์ที่ 25 ของคะแนนสอบเท่ากับ 35 คะแนน และมีนักเรียน 30 คน ได้คะแนนน้อยกว่าหรือเท่ากับ 80 คะแนน ถ้ามีนักเรียนที่สอบได้ 35 คะแนนเพียงคนเดียว แล้ว จำนวนนักเรียนที่สอบได้คะแนนในช่วง 35 - 80 คะแนน เท่ากับกี่คน [O-NET 49/1-15]

การวัดตำแหน่งข้อมูล (อันตรภาคชั้นเป็นช่วง)

ในกรณีที่ข้อมูลมาในรูปตารางที่มีอันตรภาคชั้นเป็นช่วง เราจะคำนวณ ควอไทล์ เดไซล์ เปอร์เซ็นไทล์ ยากขึ้นเยอะเลย

คะแนนสอบ	จำนวนนักเรียน (f)
1 - 10	8
11 - 20	10
21 - 30	20
31 - 40	12

ในกรณีนี้ ถ้าจะหา ควอไทล์ เดไซล์ เปอร์เซ็นไทล์ จะมีขั้นตอนคล้ายๆตอนหา Med ดังนี้

1. สร้างช่อง "ความถี่สะสม" เพื่อความสะดวกในการหาค่าข้อมูล ณ ตำแหน่งต่างๆ
2. หา "ตำแหน่ง" ด้วยสูตรใหม่

$$\boxed{\mathrm{Q}_r \rightarrow \frac{r}{4}\cdot(N)} \qquad \boxed{\mathrm{D}_r \rightarrow \frac{r}{10}\cdot(N)} \qquad \boxed{\mathrm{P}_r \rightarrow \frac{r}{100}\cdot(N)}$$

สูตรนี้ ใช้กับกรณี อันตรภาคชั้นเป็นช่วงเท่านั้น

ถ้ามาเป็นตัวๆ ให้ใช้ $N+1$ เหมือนเก่า

3. ใช้ช่องความถี่สะสม หาว่าตำแหน่งที่ต้องการ อยู่ชั้นไหน แล้วใช้สูตรไหนก็ได้ จาก 2 สูตรต่อไปนี้

$$\text{ค่าข้อมูล} = \underbrace{L}_{\text{ขอบล่าง}} + \left(\frac{\text{ตำแหน่ง} - \overbrace{F_L}^{\text{ความถี่สะสมชั้นก่อนหน้า}}}{\underbrace{f_x}_{\text{ความถี่ในชั้นนั้น}}}\right)\overbrace{I}^{\text{ความกว้างชั้น}}$$

สูตรนี้ จะเริ่มจากขอบล่าง

แล้วหาว่าต้องเดินหน้าต่อไปอีกเป็นสัดส่วนเท่าไหร่ของชั้น

$$\text{สัดส่วนที่ล้ำ} = \frac{\text{ข้อมูลล้ำไปในชั้น กี่ตัว}}{\text{จำนวนตัวในชั้น}}$$

$$\text{ค่าข้อมูล} = \underbrace{U}_{\text{ขอบบน}} - \left(\frac{\overbrace{F_x}^{\text{ความถี่สะสมในชั้นนั้น}} - \text{ตำแหน่ง}}{\underbrace{f_x}_{\text{ความถี่ในชั้นนั้น}}}\right)\overbrace{I}^{\text{ความกว้างชั้น}}$$

สูตรนี้ จะเริ่มจากขอบบน

แล้วหาว่าต้องถอยกลับมาเป็นสัดส่วนเท่าไหร่ของชั้น

$$\text{สัดส่วนที่ถอย} = \frac{\text{อีกกี่ตัว ข้อมูลถึงจะเลื่อนไปเต็มชั้น}}{\text{จำนวนตัวในชั้น}}$$

ตัวอย่าง จงหา Q_2 D_9 และ $P_{6.5}$ ของข้อมูลต่อไปนี้

คะแนนสอบ	จำนวนนักเรียน (f)	ความถี่สะสม (F)
1 - 10	8	8
11 - 20	10	18
21 - 30	20	38
31 - 40	12	50

วิธีทำ จากช่องความถี่สะสมแถวสุดท้าย จะได้ $N = 50$

Q_2 อยู่ตัวที่ $\frac{2}{4} \times 50 = 25$ $\rightarrow$ ชั้นที่ 3

$Q_2 = 20.5 + \left(\frac{25-18}{20}\right)(10)$ หรือถ้าใช้อีกสูตร $Q_2 = 30.5 - \left(\frac{38-25}{20}\right)(10)$

$= 20.5 + 3.5$ $\qquad = 30.5 - 6.5$

$= 24$ $\qquad = 24$ เท่ากัน

D_9 อยู่ตัวที่ $\frac{9}{10} \times 50 = 45$ $\rightarrow$ ชั้นสุดท้าย

$D_9 = 30.5 + \left(\frac{45-38}{12}\right)(10)$ หรือถ้าใช้อีกสูตร $D_9 = 40.5 - \left(\frac{50-45}{12}\right)(10)$

$= 30.5 + 5.83$ $\qquad = 40.5 - 4.17$

$= 36.33$ $\qquad = 36.33$ เท่ากัน

$P_{6.5}$ อยู่ตัวที่ $\frac{6.5}{100} \times 50 = 3.25$ $\rightarrow$ ชั้นแรก

$P_{6.5} = 0.5 + \left(\frac{3.25-0}{8}\right)(10)$ หรือถ้าใช้อีกสูตร $P_{6.5} = 10.5 - \left(\frac{8-3.25}{8}\right)(10)$

$= 0.5 + 4.06$ $\qquad = 10.5 - 5.94$

$= 4.56$ $\qquad = 4.56$ เท่ากัน #

ตัวอย่าง จากข้อมูลคะแนนสอบต่อไปนี้ จงหาว่า นักเรียนที่ได้คะแนนมากที่สุดในกลุ่ม 10% ต่ำสุด ได้คะแนนน้อยกว่านักเรียนที่ได้คะแนนน้อยที่สุดในกลุ่ม 10% สูงสุด อยู่กี่คะแนน

คะแนนสอบ	จำนวนนักเรียน (f)	ความถี่สะสม (F)
51 - 59	6	6
60 - 68	10	16
69 - 77	17	33
78 - 86	15	48
87 - 95	12	60

วิธีทำ คะแนนมากที่สุดในกลุ่ม 10% ต่ำสุด $\rightarrow$ P_{10}

คะแนนน้อยที่สุดในกลุ่ม 10% สูงสุด $\rightarrow$ P_{90}

และต้องใช้ความกว้างชั้น $= 60 - 51 = 9$

P_{10} อยู่ตัวที่ $\frac{10}{100} \times 60 = 6$ $\rightarrow$ ชั้นแรก ดังนั้น $P_{10} = 50.5 + \left(\frac{6-0}{6}\right)(9) = 59.5$

P_{90} อยู่ตัวที่ $\frac{90}{100} \times 60 = 54$ $\rightarrow$ ชั้นสุดท้าย ดังนั้น $P_{90} = 86.5 + \left(\frac{54-48}{12}\right)(9) = 91$

ดังนั้น P_{10} น้อยกว่า P_{90} อยู่ $= 91 - 59.5 = 31.5$ #

<u>แบบฝึกหัด</u>

1. จงหาค่าในแต่ละข้อต่อไปนี้ เมื่อกำหนดข้อมูลให้ดังนี้

คะแนนสอบ	จำนวนนักเรียน (f)
1 - 10	8
11 - 20	10
21 - 30	20
31 - 40	12

1. Med
2. Q_3
3. D_6
4. P_{85}

2. จงหาค่าในแต่ละข้อต่อไปนี้ เมื่อกำหนดข้อมูลให้ดังนี้

คะแนนสอบ	จำนวนนักเรียน (f)
1 - 5	8
6 - 10	12
11 - 15	15
16 - 20	5

1. Med
2. Q_3
3. D_1
4. $P_{90.5}$

3. กำหนดข้อมูลให้ดังตาราง จงหาว่าค่าข้อมูลต่อไปนี้ เป็นเปอร์เซ็นไทล์ที่เท่าไร

คะแนนสอบ	จำนวนนักเรียน (f)
1 - 10	5
11 - 20	12
21 - 30	20
31 - 40	3

1. 29
2. 10.1
3. 20.5
4. 26.5

การวัดการกระจายสัมบูรณ์

อีกหลายๆหัวข้อต่อจากนี้ จะสนใจว่าข้อมูลในกลุ่ม มีความแตกต่างกันมากน้อยขนาดไหน

- ถ้า ข้อมูลในกลุ่ม มีความแตกต่างกันมาก → ข้อมูลกระจายมาก
- ถ้า ข้อมูลในกลุ่ม มีความแตกต่างกันน้อย → ข้อมูลกระจายน้อย

คะแนนสอบห้อง 5/1 (ข้อมูลกระจาย<u>มาก</u>)

30	95	21	80	5
20	85	15	75	12
51	84	62	7	11

คะแนนสอบห้อง 5/2 (ข้อมูลกระจาย<u>น้อย</u>)

72	81	80	76	75
80	75	75	72	71
81	74	82	77	81

วิธีวัดการกระจายของข้อมูล จะมีหลายวิธี ได้แก่

1. พิสัย (Range)
2. ส่วนเบี่ยงเบนควอไทล์ (Quatile Deviation: Q.D.)
3. ส่วนเบี่ยงเบนเฉลี่ย (Mean Deviation: M.D.)
4. ส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation: S.D.)

ในคณิตศาสตร์พื้นฐานนี้ จะพูดถึงเฉพาะ พิสัย และ ส่วนเบี่ยงเบนมาตรฐานเท่านั้น

พิสัย

พิสัย (Range) เป็นวิธีที่วัดการกระจายที่ง่ายที่สุด
ซึ่งหาได้จากสูตร

$$\text{พิสัย} = \text{ข้อมูลมากสุด} - \text{ข้อมูลน้อยสุด}$$

ตัวอย่าง จงหาพิสัยของคะแนนสอบห้อง 5/1 และ ห้อง 5/2 ต่อไปนี้

คะแนนสอบห้อง 5/1

30	95	21	80	5
20	85	15	75	12
51	84	62	7	11

คะแนนสอบห้อง 5/2

72	81	80	76	75
80	75	75	72	71
81	74	82	77	81

วิธีทำ ห้อง 5/1: คะแนนมากสุด $= 95$ คะแนนสอบน้อยสุด $= 5$
ดังนั้น พิสัย $= 95 - 5 = 90$

ห้อง 5/2: คะแนนมากสุด $= 82$ คะแนนสอบน้อยสุด $= 71$
ดังนั้น พิสัย $= 82 - 71 = 11$ #

พิสัย จะเป็นวิธีวัดการกระจายแบบหยาบๆ คิดง่าย แต่ไม่ค่อยแม่น
เพราะ พิสัย หาจากข้อมูลแค่ 2 ตัว (คือข้อมูลมากสุด กับ ข้อมูลน้อยสุด) ข้อมูลตัวอื่นจะไม่มีผลกับพิสัย
เช่น $2\quad 7\quad 7\quad 8\quad 8\quad 9$ → พิสัย $= 9 - 2 = 7$
$2\quad 3\quad 4\quad 5\quad 7\quad 9$ → น่าจะกระจายมากกว่า แต่ พิสัย $= 9 - 2 = 7$ เท่ากัน
ดังนั้น เราจึงไม่ค่อยนิยมใช้พิสัยในการวัดการกระจาย

แบบฝึกหัด

1. จงหาพิสัยของข้อมูลต่อไปนี้

1.

5	12	10	22
11	20	16	8
14	29	20	6
16	12	9	18
18	15	34	34

2.

8	6	2	0	3	
9	8	3	1	2	7
10	4	9	3		
11	2	3			

2. คะแนนสอบของนักเรียน 3 คน มีพิสัยเท่ากับ 8 คะแนน ถ้ามัธยฐานคือ 12 คะแนน และ ค่าเฉลี่ยเลขคณิตคือ 14 คะแนน แล้ว จงหาคะแนนของนักเรียนทั้ง 3 คนนี้

3. ข้อมูลชุดหนึ่งเรียงจากน้อยไปมากดังนี้ a 11 15 18 25 b 36 41 47 53
ถ้าข้อมูลชุดนี้มีมัธยฐานเท่ากับ 28 และค่าเฉลี่ยเลขคณิตเท่ากับ 28.5 แล้ว พิสัยของข้อมูลชุดนี้เท่ากับเท่าใด [O-NET 58/39]

4. ข้อมูลชุดหนึ่งมี 8 ค่า เรียงจากน้อยไปมาก ดังนี้ 74 78 80 80 a 90 90 b
ถ้าข้อมูลชุดนี้มีพิสัยเท่ากับ 18 และมัธยฐานเท่ากับ 85 แล้วค่าเฉลี่ยเลขคณิต เท่ากับเท่าใด [O-NET 59/38]

5. ความสูงในหน่วยเซนติเมตรของนักเรียนกลุ่มหนึ่งซึ่งมี 10 คน เป็นดังนี้
155, 157, 158, 158, 160, 161, 161, 163, 165, 166
ถ้ามีนักเรียนเพิ่มขึ้นอีกหนึ่งคน ซึ่งมีความสูง 158 เซนติเมตร แล้ว ค่าสถิติใดต่อไปนี้<u>ไม่เปลี่ยนแปลง</u> [O-NET 52/28]
   1. ค่าเฉลี่ยเลขคณิต
   2. มัธยฐาน
   3. ฐานนิยม
   4. พิสัย

6. ส่วนสูงของพี่น้อง 2 คน มีพิสัยเท่ากับ 12 เซนติเมตร มีค่าเฉลี่ยเลขคณิตเท่ากับ 171 เซนติเมตร ข้อใดต่อไปนี้เป็นส่วนสูงของพี่หรือน้องคนใดคนหนึ่ง [O-NET 52/26]
   1. 167 เซนติเมตร
   2. 172 เซนติเมตร
   3. 175 เซนติเมตร
   4. 177 เซนติเมตร

7. ครอบครัวหนึ่งมีบุตร 4 คน บุตร 2 คนมีน้ำหนักเท่ากันและมีน้ำหนักน้อยกว่าบุตรอีก 2 คน
ถ้าน้ำหนักของบุตรทั้ง 4 คนมีค่าฐานนิยม มัธยฐาน และพิสัย เท่ากับ 45, 47.5 และ 7 กิโลกรัมตามลำดับ แล้ว
ค่าเฉลี่ยเลขคณิตของน้ำหนักของบุตรทั้ง 4 คน มีค่าเท่ากับกี่กิโลกรัม [O-NET 49/1-28]

8. สำหรับข้อมูลเชิงปริมาณใดๆ ที่มีค่าสถิติต่อไปนี้ ค่าสถิติใดจะตรงกับค่าของข้อมูลค่าหนึ่งเสมอ [O-NET 53/30]

1. พิสัย
2. ค่าเฉลี่ยเลขคณิต
3. มัธยฐาน
4. ฐานนิยม

ส่วนเบี่ยงเบนมาตรฐาน

ส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation: S.D.) หรือเรียกสั้นๆว่า s

สูตรสำหรับหา s จะมี 2 สูตร คือ

$$\boxed{s = \sqrt{\frac{\sum(x_i-\bar{x})^2}{N}}} \qquad \boxed{s = \sqrt{\frac{\sum x_i^2}{N} - \bar{x}^2}}$$

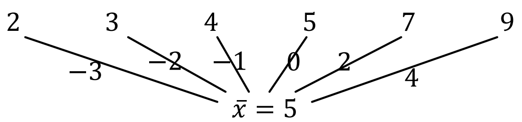
ใช้สูตรไหนก็ได้
ได้ผลลัพธ์เท่ากัน

สูตรนี้ อ่านสูตรเข้าใจยากนิดหน่อย ดูจากตัวอย่างจะเข้าใจกว่า

เช่น ถ้าต้องการหาส่วนเบี่ยงเบนสามาตรฐานของข้อมูล $2\ ,\ 3\ ,\ 4\ ,\ 5\ ,\ 7\ ,\ 9$

ถ้าหาด้วยสูตร $\sqrt{\frac{\sum(x_i-\bar{x})^2}{N}}$ จะมีขั้นตอนดังนี้

1. หา $\bar{x}$ จะได้ $\bar{x} = \frac{2+3+4+5+7+9}{6} = 5$
2. หาผลต่างของข้อมูลแต่ละตัว กับ $\bar{x}$ แล้วยกกำลังสอง ได้เป็น $(x_i - \bar{x})^2$



(ผลต่างของข้อมูลแต่ละตัวกับ $\bar{x})^2$ คือ $9\ ,\ 4\ ,\ 1\ ,\ 0\ ,\ 4\ ,\ 16$

3. เอาผลในข้อ 2 มาหาค่าเฉลี่ย แล้วถอดรูท ตอบ

$\rightarrow\ \frac{9+4+1+0+4+16}{6} = \frac{34}{6} = \frac{17}{3}$

ดังนั้น $s = \sqrt{\frac{17}{3}} = \frac{\sqrt{17}}{\sqrt{3}} \times \frac{\sqrt{3}}{\sqrt{3}} = \frac{\sqrt{51}}{3}$

หรือถ้าจะหาด้วยสูตร $\sqrt{\frac{\sum x_i^2}{N} - \bar{x}^2}$ จะมีขั้นตอนดังนี้

1. หา $\bar{x}$ จะได้ $\bar{x} = \frac{2+3+4+5+7+9}{6} = 5$
2. หา x_i^2 โดยเอาข้อมูลแต่ละตัวมายกกำลังสอง แล้วหาค่าเฉลี่ยของผลยกกำลังสอง ได้เป็น $\frac{\sum x_i^2}{N}$

$$\begin{array}{cccccc} 2 & 3 & 4 & 5 & 7 & 9 \\ \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\ 4 & 9 & 16 & 25 & 49 & 81 \end{array}$$

$\frac{\sum x_i^2}{N} = \frac{4+9+16+25+49+81}{6} = \frac{184}{6} = \frac{92}{3}$

3. เอาผลจากข้อ 2 มาลบ $\bar{x}^2$ แล้วถอดรูท ตอบ

$\rightarrow\ \frac{92}{3} - 5^2 = \frac{92}{3} - 25 = \frac{92-75}{3} = \frac{17}{3}$

ดังนั้น $s = \sqrt{\frac{17}{3}} = \frac{\sqrt{17}}{\sqrt{3}} \times \frac{\sqrt{3}}{\sqrt{3}} = \frac{\sqrt{51}}{3}$

จะเห็นว่า ค่า s มักติด $\sqrt{\ }$ ทำให้ได้ตัวเลขไม่ลงตัว อ้างอิงลำบาก
เราจะใช้คำว่า "ความแปรปรวน" (Variance) หรือ v ในการเรียก s แบบที่ไม่ต้องถอด $\sqrt{\ }$
ดังนั้น ข้อมูล 2 , 3 , 4 , 5 , 7 , 9 จะมี ความแปรปรวน $(v) = \frac{17}{3}$

อีกอย่างที่ควรทราบคือ ในกรณีที่ต้องการหา s จากกลุ่มตัวอย่าง เราจะใช้สูตร

$$s = \sqrt{\frac{\sum(x_i-\bar{x})^2}{N-1}}$$
$$s = \sqrt{\frac{\sum x_i^2 - N\bar{x}^2}{N-1}}$$

สูตรอีกชุด ที่ไม่จำเป็นต้องรู้ แต่ถ้ารู้จะทำให้ช่วยทำข้อสอบบางข้อได้เร็วขึ้น คือ สูตร "ความแปรปรวนรวม"
ถ้ามีข้อมูล 2 ชุด ชุดแรก มีข้อมูล N_1 ตัว มีค่าเฉลี่ย $= \bar{x}_1$ และมีความแปรปรวน $= S_1^2$
ชุดที่สอง มีข้อมูล N_2 ตัว มีค่าเฉลี่ย $= \bar{x}_2$ และมีความแปรปรวน $= S_2^2$
ถ้ารวมข้อมูลทั้ง 2 ชุด เป็นข้อมูลชุดใหญ่ชุดเดียว จะได้จำนวนข้อมูลทั้งหมด $N_1 + N_2$ ตัว
ถ้า $\bar{x}_1 = \bar{x}_2$ แล้ว จะได้ ค่าเฉลี่ยรวม $\bar{x}_{\text{รวม}} = \bar{x}_1 = \bar{x}_2$
ความแปรปรวนรวม $S_{\text{รวม}}^2 = \frac{N_1S_1^2+N_2S_2^2}{N_1+N_2}$
แต่ถ้า $\bar{x}_1 \neq \bar{x}_2$ แล้ว จะได้ ค่าเฉลี่ยรวม $\bar{x}_{\text{รวม}} = \frac{N_1\bar{x}_1+N_2\bar{x}_2}{N_1+N_2}$
ความแปรปรวนรวม $S_{\text{รวม}}^2 = \frac{N_1\left(S_1^2+(\bar{x}_1-\bar{x}_{\text{รวม}})^2\right)+N_2\left(S_2^2+(\bar{x}_2-\bar{x}_{\text{รวม}})^2\right)}{N_1+N_2}$

<u>ตัวอย่าง</u> นักเรียนห้องหนึ่ง เป็นชาย 16 คน เป็นหญิง 20 คน ถ้ากลุ่มนักเรียนชายมีคะแนนเฉลี่ย 22 คะแนน มีส่วนเบี่ยงเบนมาตรฐาน 2 คะแนน และกลุ่มนักเรียนหญิงมีคะแนนเฉลี่ย 22 คะแนน มีส่วนเบี่ยงเบนมาตรฐาน 3 คะแนน จงหาส่วนเบี่ยงเบนมาตรฐานของนักเรียนห้องนี้

<u>วิธีทำ</u> เนื่องจากกลุ่มนักเรียนชายและหญิง มีค่าเฉลี่ยเท่ากัน (= 22 คะแนน)
ดังนั้น จะใช้สูตรความแปรปรวนรวม $S_{\text{รวม}}^2 = \frac{N_1S_1^2+N_2S_2^2}{N_1+N_2}$ ได้
จะได้ $S_{\text{รวม}}^2 = \frac{N_{\text{ช}}S_{\text{ช}}^2+N_{\text{ญ}}S_{\text{ญ}}^2}{N_{\text{ช}}+N_{\text{ญ}}} = \frac{(16)(2^2)+(20)(3^2)}{16+20} = \frac{64+180}{36} = \frac{244}{36} = \frac{61}{9}$
ถอดรูท จะได้ $S_{\text{รวม}} = \sqrt{\frac{61}{9}} = \frac{\sqrt{61}}{3}$ #

<u>ตัวอย่าง</u> นักเรียนห้องหนึ่ง เป็นชาย 16 คน เป็นหญิง 20 คน ถ้ากลุ่มนักเรียนชายมีคะแนนเฉลี่ย 15 คะแนน มีส่วนเบี่ยงเบนมาตรฐาน 2 คะแนน และกลุ่มนักเรียนหญิงมีคะแนนเฉลี่ย 24 คะแนน มีส่วนเบี่ยงเบนมาตรฐาน 3 คะแนน จงหาส่วนเบี่ยงเบนมาตรฐานของนักเรียนห้องนี้

<u>วิธีทำ</u> เนื่องจากกลุ่มนักเรียนชายและหญิง มีค่าเฉลี่ยไม่เท่ากัน (15 และ 24 คะแนน)
ดังนั้น ต้องหา $\bar{x}_{\text{รวม}}$ และใช้สูตรความแปรปรวนรวม $S_{\text{รวม}}^2 = \frac{N_1\left(S_1^2+(\bar{x}_1-\bar{x}_{\text{รวม}})^2\right)+N_2\left(S_2^2+(\bar{x}_2-\bar{x}_{\text{รวม}})^2\right)}{N_1+N_2}$
จะได้ $\bar{x}_{\text{รวม}} = \frac{N_1\bar{x}_1+N_2\bar{x}_2}{N_1+N_2} = \frac{16(15)+20(24)}{16+20} = \frac{240+480}{36} = \frac{720}{36} = 20$
จะได้ $S_{\text{รวม}}^2 = \frac{N_1\left(S_1^2+(\bar{x}_1-\bar{x}_{\text{รวม}})^2\right)+N_2\left(S_2^2+(\bar{x}_2-\bar{x}_{\text{รวม}})^2\right)}{N_1+N_2}$
$= \frac{(16)\left(2^2+(15-20)^2\right)+(20)\left(3^2+(24-20)^2\right)}{16+20} = \frac{(16)(29)+(20)(25)}{36} = \frac{964}{36} = \frac{241}{9}$
ถอดรูท จะได้ $S_{\text{รวม}} = \sqrt{\frac{241}{9}} = \frac{\sqrt{241}}{3}$ #

<u>แบบฝึกหัด</u>

1. จงหาส่วนเบี่ยงเบนมาตรฐาน และ ความแปรปรวน ของข้อมูลต่อไปนี้

   1. $2 , 3 , 4 , 5 , 6$
   2. $3 , 6 , 9 , 10$

2. ข้อมูลชุดหนึ่ง มีจำนวน 12 ตัว ถ้า $\sum_{i=1}^{12} (x_i - \bar{x})^2 = 192$ แล้ว จงหาส่วนเบี่ยงเบนมาตรฐานของข้อมูลชุดนี้

3. ในการสุ่มตัวอย่างเพื่อสำรวจข้อมูลราคามะนาว (ต่อผล) จากตลาด 5 แห่ง ได้ข้อมูลดังนี้

   2 10 6 8 9 (หน่วย : บาท)

   ถ้า $\bar{x}$ คือ ค่าเฉลี่ยเลขคณิตของข้อมูล s คือ ส่วนเบี่ยงเบนมาตรฐานของข้อมูล
   แล้ว ร้อยละของจำนวนข้อมูลที่อยู่ในช่วง $(\bar{x} - s , \bar{x} + s)$ เท่ากับเท่าใด
   (กำหนดให้ $\sqrt{2} = 1.41$, $\sqrt{2.5} = 1.58$, $\sqrt{10} = 3.16$) [O-NET 59/39]

4. <u>กลุ่มตัวอย่าง</u>ของข้อมูลชุดหนึ่ง มี 11 จำนวนดังนี้ $15\ ,\ 10\ ,\ 12\ ,\ 15\ ,\ 16\ ,\ x\ ,\ 16\ ,\ 19\ ,\ 13\ ,\ 17\ ,\ 15$
   ถ้าค่าเฉลี่ยเลขคณิตของข้อมูลชุดนี้เท่ากับ 15
   แล้ว กำลังสองของส่วนเบี่ยงเบนมาตรฐานของข้อมูลชุดนี้เท่ากับเท่าใด [O-NET 56/31*]

5. กำหนดให้ y เป็นรายได้ต่อเดือนของพนักงาน (หน่วย : หมื่นบาท)
   และ x เป็นจำนวนปีที่พนักงานใช้ในการศึกษาระดับอุดมศึกษา
   โดย x และ y สัมพันธ์กัน ดังนี้ $y_i = 2x_i + 1$ $\quad i = 1, 2, \ldots$
   ถ้าพนักงานสี่คน ซึ่งมีรายได้ต่อเดือนเป็น $5\ ,\ 7\ ,\ 9\ ,\ a$ (หมื่นบาท)
   และค่าเฉลี่ยเลขคณิต $(\bar{x})$ ของจำนวนปีที่พนักงานใช้ในการศึกษาระดับอุดมศึกษาเท่ากับ 4 แล้ว ความแปรปรวนของรายได้ต่อเดือน เท่ากับเท่าใด [O-NET 59/30]

6. ข้อมูลชุดหนึ่งเรียงลำดับจากน้อยไปมากดังนี้
   $2 \quad 3 \quad 3 \quad x \quad 4 \quad y \quad 7$
   ถ้าค่าเฉลี่ยเลขคณิตและส่วนเบี่ยงเบนมาตรฐานของข้อมูลชุดนี้เท่ากับ 4 และ $\frac{4}{\sqrt{7}}$ ตามลำดับ
   แล้ว $y - x$ มีค่าเท่าใด [O-NET 54/37]

7. มีข้อมูล 5 จำนวนซึ่งเรียงจากน้อยไปหามาก คือ $x_1,\ x_2,\ x_3,\ x_4,\ x_5$ โดยมี $x_1 = 7$ ค่าเฉลี่ยเลขคณิตเท่ากับ $\overline{x}$ และความแปรปรวนเท่ากับ 16 ถ้ากำหนดตารางแสดงค่าของ $x_i - \overline{x}$ ดังนี้

i	$x_i - \overline{x}$
1	$7 - \overline{x}$
2	-3
3	-1
4	3
5	6

แล้ว ค่าของ $\overline{x}$ เท่ากับเท่าใด [O-NET 49/1-31]

8. โรงเรียนแห่งหนึ่งมีชั้น ม.6 อยู่สองห้องคือ 6/1 และ 6/2 ซึ่งมีจำนวนนักเรียน 52 และ 48 คน ตามลำดับ ถ้าคะแนนสอบของนักเรียน ม.6 ทั้งสองห้องมีค่าเฉลี่ยเลขคณิตเท่ากัน และมีส่วนเบี่ยงเบนมาตรฐานเท่ากับ 2 และ 1.5 ตามลำดับ แล้วส่วนเบี่ยงเบนมาตรฐานของคะแนนสอบของชั้น ม.6 เท่ากับเท่าใด [O-NET 58/32]

1. $\sqrt{3.12}$ 2. $\sqrt{3.14}$ 3. $\sqrt{3.16}$ 4. 1.75 5. 1.76

9. กำหนดให้ข้อมูลชุดที่หนึ่งซึ่งประกอบด้วย $x_1,\ x_2, \ldots,\ x_{10}$ มีค่าเฉลี่ยเลขคณิตเท่ากับ $\overline{x}$ และ ข้อมูลชุดที่สองซึ่งประกอบด้วย $y_1,\ y_2, \ldots,\ y_{20}$ มีค่าเฉลี่ยเลขคณิตเท่ากับ $\overline{y}$ โดยที่ $\sum_{i=1}^{10} (x_i - \overline{x})^2 = 160$, $\sum_{i=1}^{20} (y_i - \overline{y})^2 = 110$ และ $\overline{x} = \overline{y}$ ถ้านำข้อมูลทั้งสองชุดมารวมเป็นชุดเดียวกันแล้ว ส่วนเบี่ยงเบนมาตรฐานของข้อมูลชุดใหม่เท่ากับเท่าใด [O-NET 51/20]

สมบัติของส่วนเบี่ยงเบนมาตรฐาน

ส่วนเบี่ยงเบนมาตรฐาน มีสมบัติที่น่าสนใจดังนี้

- s จะไม่มีทางติดลบ ถึงแม้ข้อมูลจะเป็นเลขติดลบ ก็ไม่มีทางได้ s ติดลบ
  เพราะการยกกำลังสอง จะทำให้เลขลบกลายเป็นบวก

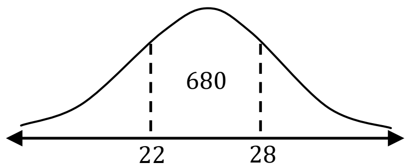
- ถ้า $s = 0$ แปลว่าข้อมูลชุดนั้น “ไม่กระจาย” เกิดขึ้นได้เมื่อข้อมูลทุกตัวเท่ากันหมด

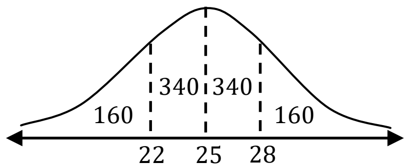
- “กระจายแบบปกติ” ของข้อมูลส่วนใหญ่ที่มีอยู่ตามธรรมชาติ มักจะ “กองอยู่ตรงกลางแบบสมมาตร” เสมอ

$$\left.\begin{array}{l}\text{จะมีข้อมูลประมาณ } 68\% \text{ อยู่ระหว่าง } \bar{x}-s \text{ กับ } \bar{x}+s \\ \text{จะมีข้อมูลประมาณ } 95\% \text{ อยู่ระหว่าง } \bar{x}-2s \text{ กับ } \bar{x}+2s \\ \text{จะมีข้อมูลประมาณ } 99\% \text{ อยู่ระหว่าง } \bar{x}-3s \text{ กับ } \bar{x}+3s\end{array}\right\} \text{ กฎ } 68 - 95 - 99$$

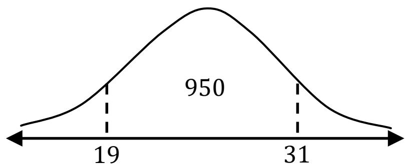
  เช่น ถ้ามีข้อมูล 1000 ตัว กระจายแบบปกติ โดยที่ $\bar{x} = 25$ และ $s = 3$

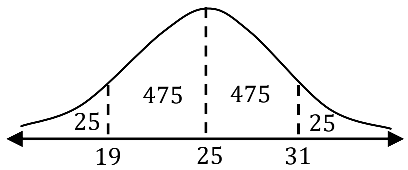
$68\% = \frac{68}{100} \times 1000$
$= 680$ คน
อยู่ระหว่าง 22 กับ 28



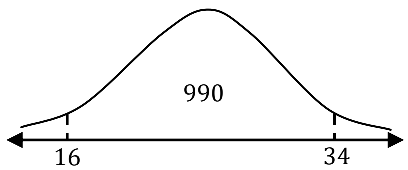


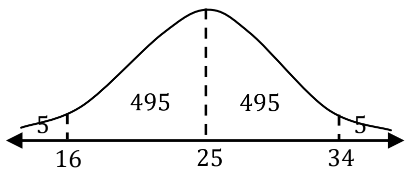
$95\% = \frac{95}{100} \times 1000$
$= 950$ คน
อยู่ระหว่าง 19 กับ 31





$99\% = \frac{99}{100} \times 1000$
$= 990$ คน
อยู่ระหว่าง 16 กับ 34





- s จะมีค่าประมาณ $\frac{\text{พิสัย}}{4}$ (สูตรนี้จะใช้เวลาต้องการประมาณค่า s แบบเร็วๆ)

- การเพิ่มหรือลดข้อมูลทุกตัวอย่างเท่าๆกัน โดยการบวกหรือลบด้วยค่าคงที่ จะไม่ทำให้ค่าการกระจายเปลี่ยน
  เช่น ถ้าเอาข้อมูล $2, 3, 4, 5, 7, 9$ มาบวก 50 เพิ่มให้ข้อมูลทุกตัว
  จะได้เป็น $52, 53, 54, 55, 57, 59$
  ข้อมูลชุดใหม่หลังบวก จะมี s เท่ากับข้อมูลชุดเดิมที่ยังไม่ได้บวก

- การเปลี่ยนข้อมูลทุกตัวอย่างเท่าๆกัน โดยการคูณด้วย k จะทำให้การกระจายเปลี่ยนเป็น $|k|$ เท่าของของเดิม
  การเปลี่ยนข้อมูลทุกตัวอย่างเท่าๆกัน โดยการหารด้วย k จะทำให้การกระจายเปลี่ยนเป็น $\frac{1}{|k|}$ เท่าของของเดิม
  เช่น ถ้าเอา $2, 3, 4, 5, 7, 9$ มาคูณ -10 ได้เป็น $-20, -30, -40, -50, -70, -90$
  ข้อมูลชุดใหม่ จะมี s เป็น $|-10| = 10$ เท่าของของเดิม

ตัวอย่าง ข้อมูลอุณหภูมิประจำวัน ในหน่วย °C ชุดหนึ่ง ประกอบด้วย 28 , 27 , 29 , 27 , 24 มีส่วนเบี่ยงเบนมาตรฐาน 2.8 เมื่อแปลงข้อมูลนี้ให้อยู่ในหน่วย °F ด้วยสูตร $\left(\mathrm{C}\times\frac{9}{5}\right)+32$ จะกลายเป็น 82.4 , 80.6 , 84.2 , 80.6 , 75.2 จงหาส่วนเบี่ยงเบนมาตรฐานของข้อมูลชุดนี้ หลังจากที่แปลงเป็นหน่วย °F แล้ว

วิธีทำ ข้อนี้ จะหา s ตรงๆ จาก 82.4 , 80.6 , 84.2 , 80.6 , 75.2 ตรงๆ ก็ได้ แต่จะใช้แรงเยอะ

จะเห็นว่าข้อมูล 28 , 27 , 29 , 27 , 24 จะถูกแปลงโดย คูณ $\frac{9}{5}$ แล้วบวกด้วย 32

การคูณด้วย $\frac{9}{5}$ จะทำให้ s เปลี่ยนไป $\frac{9}{5}$ เท่า

แต่การบวกด้วย 32 จะไม่ทำให้ s เปลี่ยน

ดังนั้น ข้อมูลหลังแปลง จะมี $s_{\text{ใหม่}} = s_{\text{เดิม}}\times\frac{9}{5} = 2.8\times\frac{9}{5} = 5.04$ #

แบบฝึกหัด

1. ช่วงก่อนสอบ น้ำหนักของนักเรียน ม. 5/1 จำนวน 40 คน มีส่วนเบี่ยงเบนมาตรฐานเท่ากับ 4 กก. เมื่อสอบเสร็จพบว่านักเรียนทุกคนน้ำหนักเพิ่มขึ้นคนละ 2 กก. จงหาส่วนเบี่ยงเบนมาตรฐานของน้ำหนักนักเรียน ม. 5/1 หลังสอบเสร็จ

2. นักเรียนกลุ่มหนึ่ง ประกอบด้วยผู้ชาย 3 คน และผู้หญิง 2 คน โดยกลุ่มผู้ชายมีความแปรปรวนเท่ากับ 0 และมีอายุเฉลี่ย 25 ปี ถ้าผู้หญิงทั้งสองคน มีอายุ 15 และ 20 ปี ตามลำดับแล้ว ความแปรปรวนของอายุนักเรียนกลุ่มนี้เท่ากับเท่าใด

3. พนักงานบริษัทจำนวน 800 คน มีเงินเดือนเฉลี่ย 10,000 บาท ส่วนเบี่ยงเบนมาตรฐาน 1,500 บาท ถ้าเงินเดือนพนักงานกลุ่มนี้ มีการกระจายแบบปกติแล้ว มีพนักงานประมาณกี่คน ที่ได้เงินเดือนน้อยกว่า 7,000 บาท

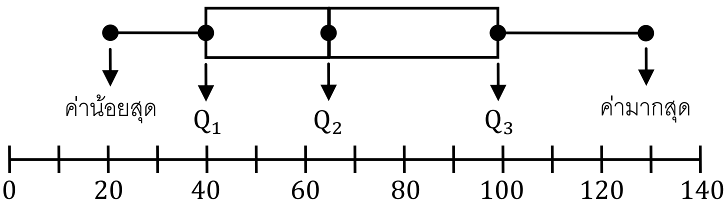
4. ข้อมูลชุดที่หนึ่งมี 10 จำนวน คือ $x_1, x_2, x_3, \dots, x_{10}$ ซึ่งข้อมูลชุดนี้มีส่วนเบี่ยงเบนมาตรฐานเท่ากับ 2.3 ถ้าข้อมูลชุดที่สองมี 10 จำนวน คือ $3x_1 + 174$, $3x_2 + 174$, $3x_3 + 174$, ... , $3x_{10} + 174$ แล้วส่วนเบี่ยงเบนมาตรฐานของข้อมูลชุดที่สองนี้จะเท่ากับเท่าใด [O-NET 57/39]

5. เมื่อสองปีก่อน นักเรียนห้องหนึ่งมี 30 คน แบ่งออกได้เป็นสองกลุ่ม กลุ่มที่หนึ่งมี 10 คน ทุกคนมีอายุ 10 ปี และกลุ่มที่สองมี 20 คน มีอายุเฉลี่ย 8.5 ปี ถ้าความแปรปรวนของอายุนักเรียนกลุ่มที่สอง เท่ากับ 0 แล้ว ในปัจจุบันความแปรปรวนของอายุนักเรียนห้องนี้ เท่ากับเท่าใด [O-NET 50/39]

6. พนักงานโรงงานแห่งหนึ่งจำนวน 1,000 คน ได้รับเงินเดือนเฉลี่ยคนละ 8,000 บาท มีส่วนเบี่ยงเบนมาตรฐาน 1,000 บาท ถ้าการกระจายของเงินเดือนพนักงานโรงงานแห่งนี้เป็นแบบปกติแล้ว ข้อสรุปในข้อใดต่อไปนี้<u>ผิด</u> [O-NET 51/40]
   1. พนักงานจำนวนน้อยกว่า 100 คน ได้รับเงินเดือนน้อยกว่า 6,000 บาท
   2. พนักงานอย่างมาก 930 คน ได้รับเงินเดือนมากกว่าหรือเท่ากับ 6,000 บาท
   3. พนักงานที่ได้รับเงินเดือนมากกว่า 10,000 บาท มีจำนวนน้อยกว่า 70 คน
   4. ถ้าในปีต่อไปพนักงานได้รับเงินเดือนเพิ่มขึ้นคนละ 400 บาทแล้ว ส่วนเบี่ยงเบนมาตรฐานของเงินเดือนพนักงานโรงงานนี้ยังคงเดิม

แผนภาพกล่อง

แผนภาพกล่อง เป็นแผนภูมิรูปภาพเพื่อบอก ค่าน้อยสุด , Q_1 , Q_2 , Q_3 และ ค่ามากสุด ดังรูป



จากแผนภาพกล่อง เราจะสามารถคำนวณ พิสัย ($=$ ค่ามากสุด $-$ ค่าน้อยสุด)

ส่วนเบี่ยงเบนควอไทล์ ($=\frac{Q_3-Q_1}{2}$)

สัมประสิทธิ์ของส่วนเบี่ยงเบนควอไทล์ ($=\frac{Q_3-Q_1}{Q_3+Q_1}$) ได้

ซึ่งค่าเหล่านี้ ใช้บอกการกระจายของข้อมูลได้

นั่นคือ ถ้าค่ามากสุด กับค่าน้อยสุด อยู่ห่างกันมาก แปลว่าข้อมูลมีการกระจายมาก

และเนื่องจาก ควอไทล์ เป็นจุดที่แบ่งข้อมูลออกเป็น 4 ส่วนเท่าๆกัน

สิ่งที่ต้องเข้าใจให้แม่น คือ "แต่ละส่วน มีข้อมูล 25% เท่ากันหมด" ไม่ขึ้นกับว่าส่วนไหนยาว ส่วนไหนสั้น

ความยาว ความสั้น จะมีผลกับ "ความแออัด" ของข้อมูล

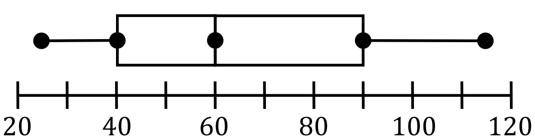
เนื่องจาก แต่ละส่วน มีข้อมูล 25% เท่ากัน ดังนั้น "ส่วนที่สั้น จะแออัด กว่าส่วนที่ยาว"

ดังนั้น เราจะบอกลักษณะ "การเบ้" ได้ด้วย โดยดูว่า "ส่วนสั้น อยู่ทางซ้ายหรือทางขวา"

ถ้า ส่วนสั้นอยู่ทางซ้าย แปลว่าทางซ้ายแออัด แปลว่า โค้งโด่งซ้าย แปลว่า ข้อมูลเบ้ขวา 

ถ้า ส่วนสั้นอยู่ทางขวา แปลว่าทางขวาแออัด แปลว่า โค้งโด่งขวา แปลว่า ข้อมูลเบ้ซ้าย 

<u>ตัวอย่าง</u> จากแผนภาพกล่องแสดงคะแนนสอบของนักเรียน จำนวน 60 คน ต่อไปนี้



จงพิจารณาว่าข้อใดผิด

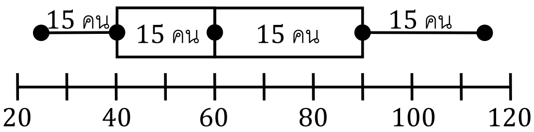
1. พิสัยของคะแนนสอบ คือ 90 คะแนน
2. ข้อมูลมีลักษณะการกระจายแบบเบ้ขวา
3. นักเรียนที่ได้คะแนนในช่วง 60 - 80 มีน้อยกว่า 15 คน
4. นักเรียนที่ได้คะแนนมากกว่า 60 มีจำนวนมากกว่านักเรียนที่ได้คะแนนน้อยกว่า 60

<u>วิธีทำ</u> จากแผนภาพ จะได้ ข้อมูลน้อยสุด $= 25$, $Q_1 = 40$, $Q_2 = 60$, $Q_3 = 90$, ข้อมูลมากสุด $= 115$

1. พิสัย $=$ ข้อมูลมากสุด $-$ ข้อมูลน้อยสุด $= 115 - 25 = 90$ ดังนั้น ข้อ 1 ถูกต้อง
2. จะเห็นว่าส่วนสั้นอยู่ทางซ้าย ดังนั้น ทางซ้ายแออัด แปลว่าโค้งโด่งซ้าย

   ดังนั้น ข้อมูลนี้เป็นแบบเบ้ขวา  ดังนั้น ข้อ 2 ก็ถูกอีก

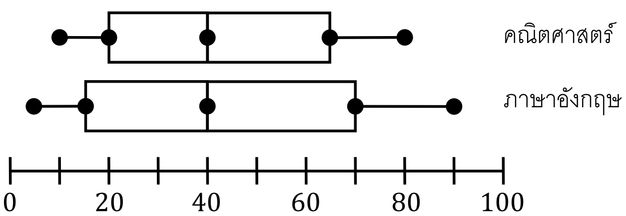
3. เนื่องจากมีนักเรียน 60 คน แบ่งเป็น 4 ส่วน ส่วนละเท่าๆกัน จะได้แต่ละส่วนจะมีนักเรียน 15 คน



จากภาพ จะได้ว่าช่วง 60 - 90 มีนักเรียน 15 คน
ดังนั้น ข้อ 3 ถูก เพราะ ช่วง 60 - 80 เล็กกว่า 60 - 90 จึงน่าจะมีนักเรียนน้อยกว่า 15 คน

4. นักเรียนที่ได้น้อยกว่า 60 คะแนน จะมี 2 ส่วน (25 - 40 และ 40 - 60) $\rightarrow$ มี $15 + 15 = 30$ คน
นักเรียนที่ได้มากกว่า 60 คะแนน ก็มี 2 ส่วน (60 - 90 และ 90 - 115) $\rightarrow$ มี $15 + 15 = 30$ คน
ดังนั้น ข้อ 4 ผิด #

<u>ตัวอย่าง</u> จากแผนภาพกล่องแสดงข้อมูลคะแนนสอบ วิชาคณิตศาสตร์ และ ภาษาอังกฤษ ของนักเรียนห้องหนึ่ง



จงพิจารณาว่าข้อใดถูก

1. คะแนนวิชาภาษาอังกฤษ มีการกระจายน้อยกว่า
2. ค่าเฉลี่ยเลขคณิตของคะแนนสอบทั้งสองวิชา เท่ากัน
3. จำนวนนักเรียนที่ได้คณิตศาสตร์ น้อยกว่า 65 คะแนน มีน้อยกว่า
   จำนวนนักเรียนที่ได้ภาษาอังกฤษ น้อยกว่า 65 คะแนน
4. คะแนนสูงสุดที่อยู่ในกลุ่ม 25 % ต่ำสุด ของวิชาคณิตศาสตร์ มากกว่า
   คะแนนสูงสุดที่อยู่ในกลุ่ม 25 % ต่ำสุด ของวิชาภาษาอังกฤษ

<u>วิธีทำ</u> 1. ภาษาอังกฤษ มีช่วงคะแนน มากสุด - น้อยสุด กว้างกว่า
ดังนั้น ภาษาอังกฤษมีการกระจายมากกว่า ข้อ 1 จึงผิด

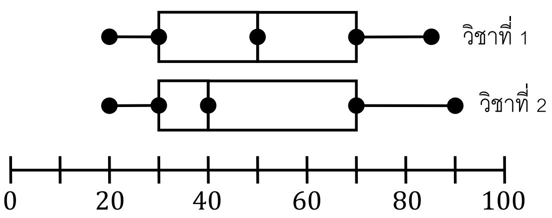
2. เนื่องจากแผนภาพกล่อง บอกแต่ควอไทล์
ข้อมูลที่โจทย์ให้ ไม่สามารถนำไปสรุปเกี่ยวกับค่าเฉลี่ยได้ ดังนั้น ข้อ 2 จะจริงหรือเปล่า ก็ไม่รู้

3. นักเรียนที่ได้คณิตศาสตร์ น้อยกว่า 65 คะแนน จะมี 3 ส่วน
แต่นักเรียนที่ได้อังกฤษ น้อยกว่า 65 คะแนน จะมีแค่ 2 ส่วนกว่าๆ ดังนั้น ข้อ 3 ผิด

4. "คะแนนสูงสุดที่อยู่ในกลุ่ม 25 % ต่ำสุด" หมายถึง Q_1 นั่นเอง
จะเห็นว่า วิชาคณิตศาสตร์ มี $Q_1 = 20$
วิชาภาษาอังกฤษ มี $Q_1 = 15$ ดังนั้น ข้อ 4 ถูกต้อง #

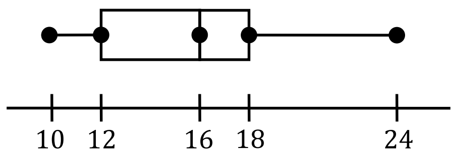
<u>แบบฝึกหัด</u>

1. จากแผนภาพกล่องแสดงคะแนนสอบ 2 วิชา ของนักเรียนห้อง ม. 5/1 ต่อไปนี้ จงพิจารณาว่าข้อใดถูกต้องบ้าง



1. Q_1 ของวิชาที่ 1 มากกว่า Q_1 ของวิชาที่ 2
2. Q_2 ของวิชาที่ 1 มากกว่า Q_2 ของวิชาที่ 2
3. Q_3 ของวิชาที่ 1 มากกว่า Q_3 ของวิชาที่ 2
4. คะแนนเฉลี่ยของวิชาที่ 1 มากกว่า คะแนนเฉลี่ยของวิชาที่ 2
5. คะแนนต่ำสุด ในกลุ่ม 25% สูงสุด ในวิชาที่ 1 มากกว่า วิชาที่ 2
6. มีนักเรียนที่สอบได้คะแนน 30 - 40 คะแนน ในวิชาที่ 1 มากกว่า วิชาที่ 2
7. มีนักเรียนที่สอบได้คะแนน 50 - 70 คะแนน ในวิชาที่ 1 มากกว่า วิชาที่ 2
8. มีนักเรียนที่สอบได้คะแนนน้อยกว่า 40 คะแนน ในวิชาที่ 1 มากกว่า วิชาที่ 2
9. มีนักเรียนที่สอบได้คะแนนน้อยกว่า 50 คะแนน ในวิชาที่ 1 มากกว่า วิชาที่ 2
10. มีนักเรียนที่สอบได้คะแนนน้อยกว่า 70 คะแนน ในวิชาที่ 1 มากกว่า วิชาที่ 2

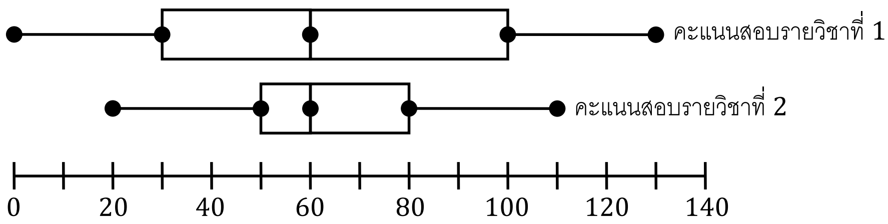
2. คะแนนสอบความรู้ทั่วไปของนักเรียน 200 คนนำเสนอโดยใช้แผนภาพกล่องดังนี้



ข้อใดเป็น<u>เท็จ</u> [O-NET 53/32]

1. จำนวนนักเรียนที่ทำได้ 12 ถึง 16 คะแนน มีเท่ากับ จำนวนนักเรียนที่ทำได้ 16 ถึง 18 คะแนน
2. จำนวนนักเรียนที่ทำได้ 12 ถึง 18 คะแนน มีเท่ากับ จำนวนนักเรียนที่ทำได้ 18 ถึง 24 คะแนน
3. จำนวนนักเรียนที่ทำได้ 10 ถึง 12 คะแนน มีเท่ากับ จำนวนนักเรียนที่ทำได้ 18 ถึง 24 คะแนน
4. จำนวนนักเรียนที่ทำได้ 10 ถึง 16 คะแนน มีเท่ากับ จำนวนนักเรียนที่ทำได้ 16 ถึง 24 คะแนน

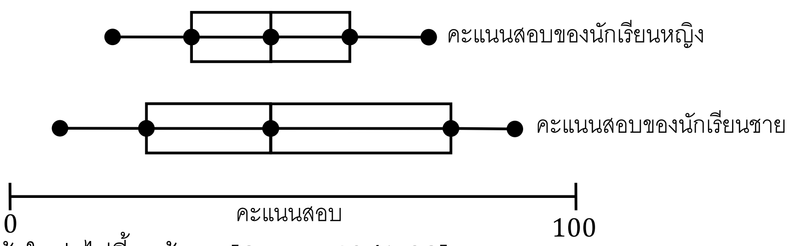
3. จากการทดสอบนักเรียนจำนวน 100 คนใน 2 รายวิชา แต่ละรายวิชามีคะแนนเต็ม 150 คะแนน ถ้าผลการทดสอบทั้งสองรายวิชา เขียนเป็นแผนภาพกล่องได้ดังนี้ [O-NET 50/40]



แล้ว ข้อสรุปในข้อใดต่อไปนี้ถูก

1. คะแนนสอบทั้งสองรายวิชามีการแจกแจงแบบปกติ
2. จำนวนนักเรียนที่ได้คะแนนไม่เกิน 80 คะแนน ในรายวิชาที่ 1 มากกว่า
   จำนวนนักเรียนที่ได้คะแนนไม่เกิน 80 คะแนน ในรายวิชาที่ 2
3. คะแนนสูงสุดที่อยู่ในกลุ่ม 25 % ต่ำสุด ของผลการสอบรายวิชาที่ 1 น้อยกว่า
   คะแนนสูงสุดที่อยู่ในกลุ่ม 25 % ต่ำสุด ของผลการสอบรายวิชาที่ 2
4. จำนวนนักเรียนที่ได้คะแนนระหว่าง 60 - 80 คะแนน ในการสอบรายวิชาที่ 2 น้อยกว่า
   จำนวนนักเรียนที่ได้คะแนนในช่วงเดียวกัน ในการสอบรายวิชาที่ 1

4. จากแผนภาพกล่องของคะแนนสอบวิชาคณิตศาสตร์ของนักเรียนจำแนกตามเพศเป็นดังนี้



ข้อสรุปในข้อใดต่อไปนี้ถูกต้อง [O-NET 49/1-32]

1. คะแนนสอบเฉลี่ยวิชาคณิตศาสตร์ของนักเรียนชาย
   สูงกว่าคะแนนสอบเฉลี่ยวิชาคณิตศาสตร์ของนักเรียนหญิง
2. คะแนนสอบวิชาคณิตศาสตร์ของนักเรียนชายมีการกระจายเบ้ขวา
3. คะแนนสอบวิชาคณิตศาสตร์ของนักเรียนหญิงมีการกระจาย
   มากกว่าคะแนนสอบวิชาคณิตศาสตร์ของนักเรียนชาย
4. คะแนนสอบวิชาคณิตศาสตร์ของนักเรียนหญิงมีการกระจายเบ้ขวา