

สถิติ ม.6

ความหมายของสถิติศาสตร์และข้อมูล

การที่จะคิดหรือตัดสินใจบางอย่าง สิ่งที่เป็นที่สุดคือ **ข้อมูล** ตั้งแต่ปริมาณน้อยจนปริมาณขนาดใหญ่ ข้อมูลทั้งหลายต่างก็มีประโยชน์ ผู้คนจึงหาวิธีการจัดการข้อมูลเหล่านี้เพื่อนำไปใช้ได้ง่ายขึ้น โดยกระบวนการหรือความรู้ต่าง ๆ จึงมารวมกัน เป็นศาสตร์ที่เรานิยมเรียกกันว่า **สถิติศาสตร์**

คำสำคัญทางสถิติ

ประชากร (population) กลุ่มของหน่วยทั้งหมดในเรื่องที่สนใจศึกษา หน่วยในที่นี้อาจเป็น คน สัตว์ หรือสิ่งของ	ตัวอย่าง (sample) กลุ่มย่อยของประชากรที่ถูกเลือกมาเป็นตัวแทนของประชากร โดยทั่วไปมีวัตถุประสงค์เพื่อใช้ตัวอย่างในการสรุปผลเกี่ยวกับลักษณะของประชากรที่สนใจ	ตัวแปร (variable) ลักษณะบางประการของประชากรหรือตัวอย่างที่สนใจศึกษา
ข้อมูล (data) ข้อความที่เกี่ยวกับเรื่องใดเรื่องหนึ่งที่สามารถใช้ในการสรุปผลในเรื่องที่สนใจศึกษา อาจเป็นได้ทั้งตัวเลขหรือไม่ใช่ตัวเลข หรืออาจหมายถึงค่าของตัวแปรที่สนใจศึกษา	พารามิเตอร์ (parameter) ค่าวัดที่แสดงลักษณะของประชากร ซึ่งเป็นค่าคงตัวที่คำนวณหรือประมวลจากข้อมูลทั้งหมดของประชากร	ค่าสถิติ (statistic) ค่าคงตัวที่พิจารณาจากข้อมูลของตัวอย่าง โดยมีวัตถุประสงค์เพื่ออธิบายลักษณะของตัวอย่างนั้น หรือเพื่อประมาณค่าของพารามิเตอร์แล้วนำไปใช้ในการอธิบายลักษณะของประชากร

ประเภทของข้อมูล

1. การแบ่งประเภทของข้อมูลตามแหล่งที่มาของข้อมูล

ข้อมูลปฐมภูมิ (primary data)

คือ ข้อมูลที่ผู้ใช้ดำเนินการเก็บรวบรวมจากเจ้าของข้อมูลหรือต้นกำเนิดของข้อมูลโดยตรง

ข้อมูลทุติยภูมิ (secondary data)

คือ ข้อมูลที่ผู้ใช้ไม่ได้ดำเนินการเก็บรวบรวมจากเจ้าของข้อมูลหรือต้นกำเนิดของข้อมูลโดยตรง แต่ใช้ข้อมูลจากบุคคลหรือหน่วยงานอื่น

2. การแบ่งประเภทของข้อมูลตามระยะเวลาที่จัดเก็บ

ข้อมูลอนุกรมเวลา (time series data)

คือ ชุดข้อมูลที่เกิดขึ้นและจัดเก็บตามลำดับเวลาต่อเนื่องกันไปตลอดช่วง ๆ หนึ่ง

ข้อมูลตัดขวาง (cross-sectional data)

คือ ข้อมูลที่บอกสถานะหรือสภาพของสิ่งที่สนใจ ณ จุดหนึ่งของเวลา

3. การแบ่งประเภทของข้อมูลตามลักษณะของข้อมูล

ข้อมูลเชิงปริมาณ (quantitative data)

คือ ข้อมูลที่ได้จากการวัดหรือการนับค่า โดยแสดงเป็นตัวเลขหรือปริมาณที่สามารถนำไปบวก ลบ คูณ หรือหาร และเปรียบเทียบกันได้

ข้อมูลเชิงคุณภาพ (qualitative data)

คือ ข้อมูลที่แสดงลักษณะ ประเภท สมบัติ ในเชิงคุณภาพ และอื่น ๆ ที่ไม่สามารถวัดค่าเป็นตัวเลขที่นำมาบวก ลบ คูณ หรือหารกันได้

ประเภทของการวิเคราะห์ข้อมูล

สถิติศาสตร์เชิงพรรณนา (descriptive statistics)

คือ การวิเคราะห์ข้อมูลที่สรุปสาระสำคัญของข้อมูลชุดหนึ่ง ซึ่งเป็นข้อมูลเชิงคุณภาพหรือข้อมูลเชิงปริมาณ เพื่ออธิบายลักษณะหรือสภาพของข้อมูลชุดนั้นว่าเป็นอย่างไ

สถิติศาสตร์เชิงอนุมาน (inferential statistics)

คือ การวิเคราะห์ข้อมูลที่ใช้ทฤษฎีที่เกี่ยวข้องกับความน่าจะเป็นในการหาข้อสรุปเกี่ยวกับลักษณะของประชากร โดยใช้ข้อมูลจากตัวอย่างที่ได้มาจากประชากรนั้น

การวิเคราะห์ข้อมูลและนำเสนอข้อมูลเชิงคุณภาพ

ในหัวข้อนี้ เราจะวิเคราะห์ข้อมูลเชิงคุณภาพโดยพิจารณาจากความถี่และฐานนิยม

ความถี่ (frequency)

คือ จำนวนครั้งของการเกิดข้อมูลข้อมูลหนึ่งและค่าของตัวแปรค่าหนึ่ง

ฐานนิยม (mode)

คือ ข้อมูลที่มีจำนวนครั้งของการเกิดซ้ำกันมากที่สุดหรือข้อมูลที่มีความถี่สูงสุดที่มากกว่า 1

ตามหนังสือสสวท. กล่าวไว้ว่า ข้อมูลบางชุดอาจไม่มีฐานนิยม เช่น ในกรณีที่ข้อมูลมีความถี่เป็น 1 เท่ากันหมด นอกจากนี้ ข้อมูลบางชุดอาจมีฐานนิยมมากกว่า 1 ค่า อย่างไรก็ตาม ในที่นี้จะพิจารณาเฉพาะชุดข้อมูลที่มีฐานนิยมเพียงค่าเดียว

การวิเคราะห์และนำเสนอข้อมูลเชิงคุณภาพด้วยตารางความถี่

การนำเสนอข้อมูลเชิงคุณภาพในรูปตารางความถี่ (frequency table) จะมีด้วยกัน 2 รูปแบบคือ ตารางความถี่จำแนกทางเดียว (one-way frequency table) และ ตารางความถี่จำแนกสองทาง (two-way frequency table)

ตารางความถี่จำแนกทางเดียว

เป็นการนำเสนอข้อมูลในรูปตารางของตัวแปรหนึ่งตัว เช่น

การสำรวจสีเสื้อผ้าของเด็ก ๆ ทั้ง 10 คน สามารถเขียนตารางความถี่ได้ดังนี้

สีเสื้อ	แดง	ดำ	ขาว	รวม
ความถี่	1	4	5	10

หรือ

สีเสื้อ	ความถี่
แดง	1
ดำ	4
ขาว	5
รวม	10

นอกจากนี้เราอาจเปรียบเทียบข้อมูลในตารางโดยใช้ความถี่สัมพัทธ์ได้เช่นกัน

ความถี่สัมพัทธ์

คือ สัดส่วนของความถี่เทียบกับความถี่ทั้งหมด ซึ่งอาจเขียนได้ 2 รูปแบบ คือ แบบสัดส่วน และ แบบร้อยละ

ความถี่สัมพัทธ์

สัดส่วน	ร้อยละ
$\text{ความถี่สัมพัทธ์ (สัดส่วน)} = \frac{\text{ความถี่}}{\text{ความถี่รวม}}$	$\text{ความถี่สัมพัทธ์ (ร้อยละ)} = \frac{\text{ความถี่}}{\text{ความถี่รวม}} \times 100$

ตัวอย่างที่ 1 จากข้อมูลตารางความถี่ของการสำรวจสีเสื้อของเด็ก ๆ ทั้ง 10 คน ที่กำหนดให้
จงเขียนแสดงความถี่สัมพัทธ์ของแต่ละข้อมูล

ตัวอย่างข้อมูลความถี่สัมพัทธ์

สีเสื้อ	ความถี่	ความถี่สัมพัทธ์	
		สัดส่วน	ร้อยละ
แดง	1	$\frac{1}{10} = 0.1$	$0.1 \times 100 = 10$
ดำ	4	$\frac{4}{10} = 0.4$	$0.4 \times 100 = 40$
ขาว	5	$\frac{5}{10} = 0.5$	$0.5 \times 100 = 50$
รวม	10	$\frac{10}{10} = 1$	$1 \times 100 = 100$

ตารางความถี่จำแนกสองทาง

เป็นการนำเสนอข้อมูลในรูปตารางซึ่งมีตัวแปรที่สนใจศึกษา 2 ตัว เรียกความถี่ของข้อมูลที่มีลักษณะ
ร่วมกันจากทั้งสองตัวแปรว่า ความถี่ร่วม เช่น

จากการสำรวจการเล่นกีฬาฟุตบอลกับกีฬาเบตมินตันของพนักงาน SMP โดยแบ่งตามเพศ สามารถ
เขียนตารางความถี่ได้ดังนี้

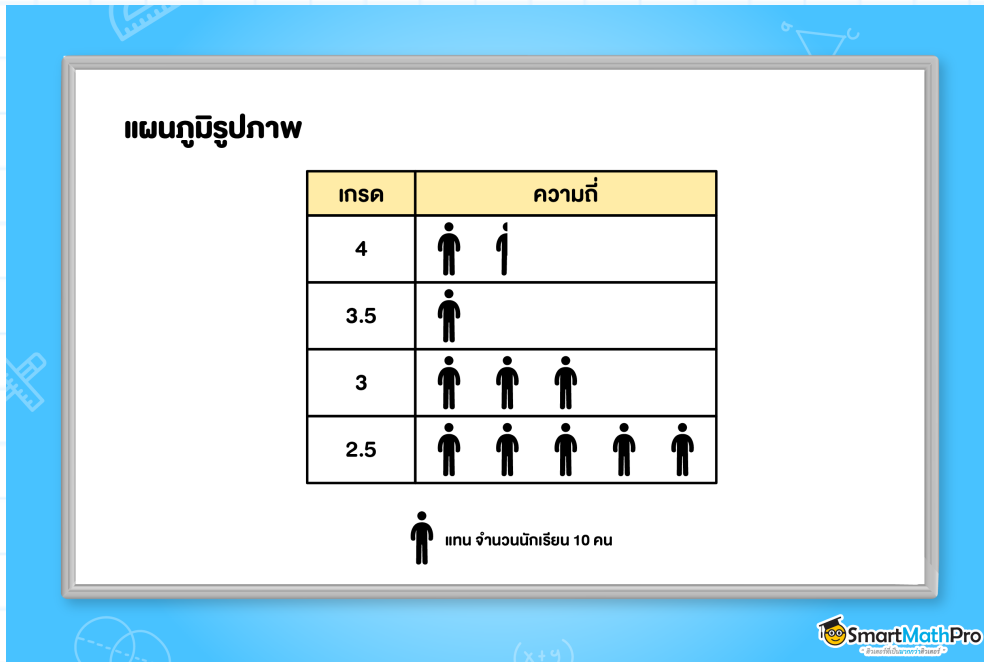
เพศ	กีฬาฟุตบอล	กีฬาเบตมินตัน	รวม
ชาย	20	25	45
หญิง	15	20	35
รวม	35	25	80

การวิเคราะห์และนำเสนอข้อมูลเชิงคุณภาพด้วยแผนภาพ

ในหัวข้อนี้จะเป็นการทบทวนแผนภูมิต่าง ๆ ที่เจอกันมาแล้วในช่วงมัธยมศึกษาตอนต้น

แผนภูมิรูปภาพ

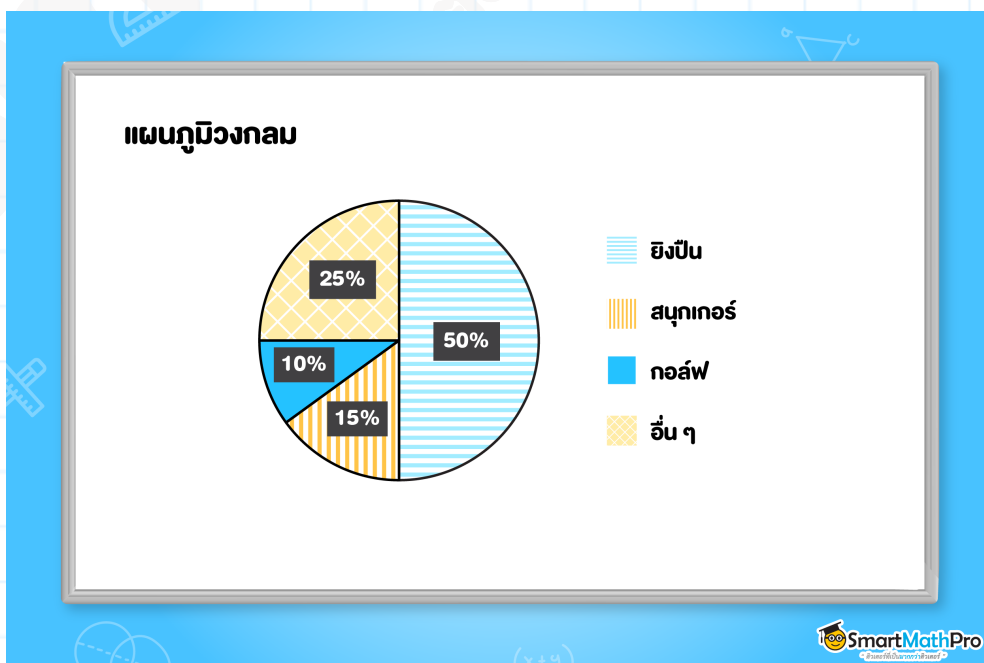
เป็นการนำเสนอข้อมูลโดยใช้รูปภาพหรือสัญลักษณ์แสดงความถี่ของแต่ละข้อมูล เช่น การสำรวจเกรดวิชาคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 สามารถเขียนแผนภูมิรูปภาพได้ดังนี้



แผนภูมิรูปวงกลม

เป็นการนำเสนอข้อมูลโดยใช้พื้นที่ภายในของรูปวงกลมแทนความถี่ โดยสัดส่วนของความถี่ของแต่ละข้อมูลเท่ากับสัดส่วนของขนาดของมุมที่จุดศูนย์กลางของรูปวงกลม เช่น

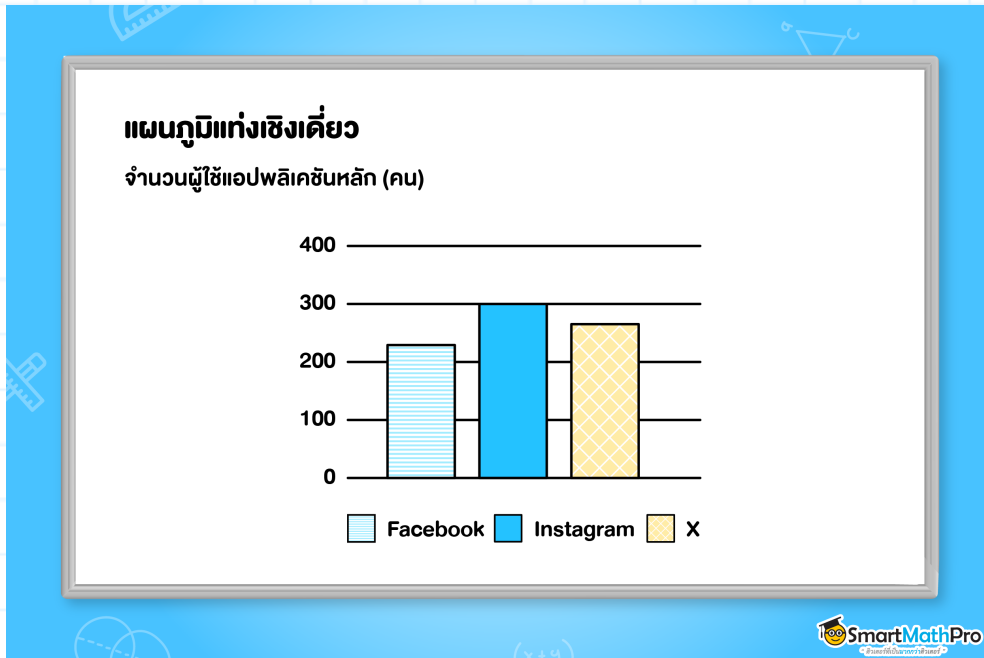
การสำรวจคนชอบกีฬาชนิดต่าง ๆ ของ พนักงานในโรงงานแห่งหนึ่ง สามารถเขียนแผนภูมิรูปวงกลมได้ดังนี้



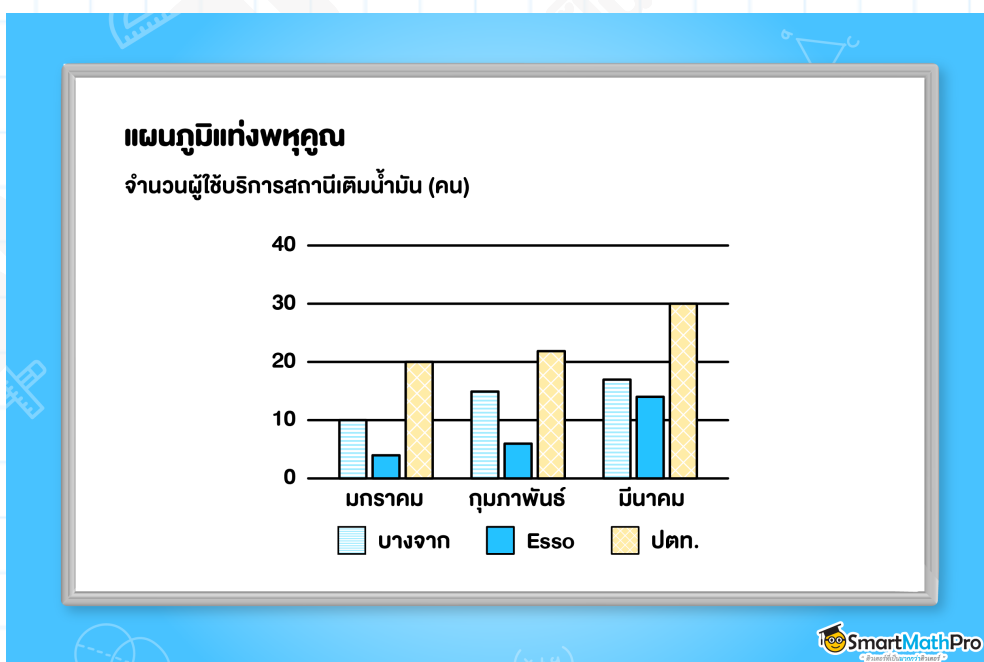
แผนภูมิแท่ง

เป็นการนำเสนอข้อมูลด้วยแท่งรูปสี่เหลี่ยมมุมฉาก โดยใช้ความยาวของแท่งแสดงความถี่ของข้อมูล มี 3 แบบ คือ แผนภูมิแท่งเชิงเดี่ยว แผนภูมิแท่งพหุคูณ และ แผนภูมิแท่งส่วนประกอบ

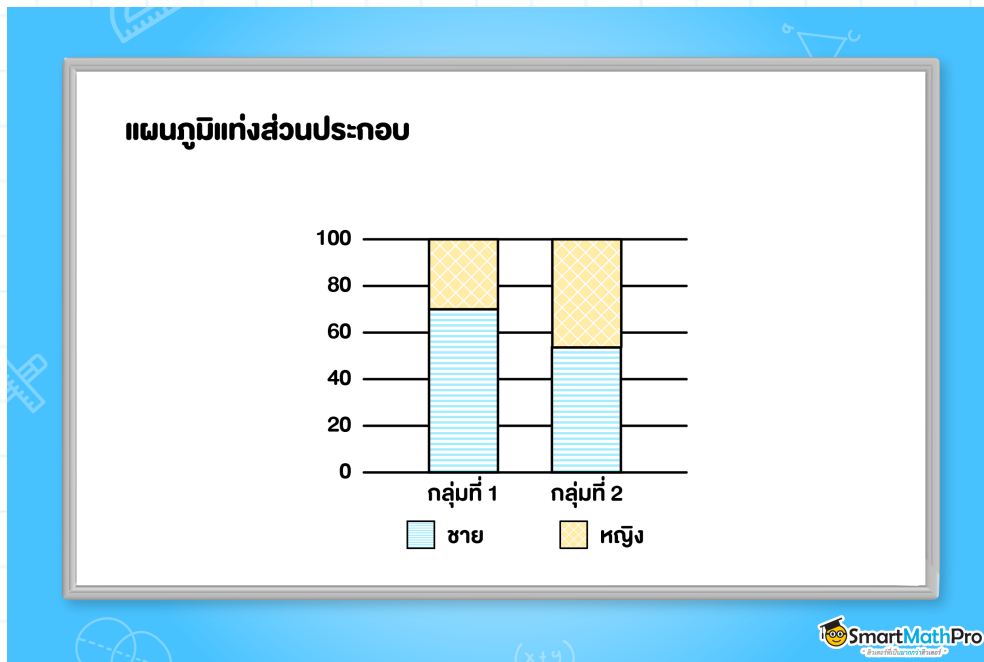
- **แผนภูมิแท่งเชิงเดี่ยว** เป็นแผนภูมิแท่งซึ่งมีตัวแปรเพียงหนึ่งตัว เช่น การสำรวจคนใช้แอปพลิเคชันต่าง ๆ สามารถเขียนแผนภูมิแท่งเชิงเดี่ยวได้ดังนี้



- **แผนภูมิแท่งพหุคูณ** เป็นแผนภูมิแท่งซึ่งมีตัวแปรที่สนใจศึกษาตั้งแต่ 2 ตัวขึ้นไป เช่น การสำรวจจำนวนผู้ใช้สถานีเติมน้ำมันต่าง ๆ ในแต่ละเดือนสามารถเขียนแผนภูมิแท่งพหุคูณได้ดังนี้



- แผนภูมิแท่งส่วนประกอบ เป็นแผนภูมิแท่งที่แสดงจำนวนรวมและส่วนประกอบของจำนวนรวมนั้น โดยการแบ่งเป็นส่วนย่อย ๆ เช่น การสำรวจผู้เข้าร่วมกิจกรรมขึ้นดอยมีจำนวนผู้เข้าร่วมแต่ละเพศที่เข้ากลุ่มซึ่งแบ่งเป็น 2 กลุ่ม กลุ่มละ 100 คน สามารถเขียนแผนภูมิแท่งส่วนประกอบได้ดังนี้



การวิเคราะห์และนำเสนอข้อมูลเชิงปริมาณด้วยตารางความถี่

จำได้รึเปล่าเอ่ยว่าข้อมูลเชิงคุณภาพที่แสดงด้วยตารางความถี่สามารถนำมาใช้ในการวิเคราะห์ต่าง ๆ ได้ ในส่วนของข้อมูลเชิงปริมาณก็สามารถทำได้เช่นกัน

ซึ่งจะใช้ตารางความถี่ 2 แบบหลัก ได้แก่ ตารางความถี่แบบไม่ได้แบ่งข้อมูลเป็นช่วง และตารางความถี่แบบแบ่งข้อมูลเป็นช่วง

- ตารางความถี่แบบไม่ได้แบ่งข้อมูลเป็นช่วง จะเหมาะสำหรับค่าความเป็นไปได้ของข้อมูลที่มีจำนวนน้อย เช่น จำนวนวัน Work from home ของพนักงานในบริษัท SMP

เวลา (วัน)	1	2	3	4	5
จำนวนคน	21	5	20	10	15

- ตารางความถี่แบบแบ่งเป็นช่วง จะเหมาะสำหรับค่าที่เป็นไปได้ของข้อมูลมีจำนวนมาก การเขียนตารางความถี่ทำได้โดยการหาความกว้างของอันตรภาคชั้นจาก

$$\frac{(\text{ค่าสุดท้าย} - \text{ค่าเริ่มต้น})}{\text{จำนวนอันตรภาคชั้นที่ต้องการ}}$$

จำไว้ว่าหากค่าที่คำนวณได้ไม่เป็นจำนวนเต็มจะต้องปัดขึ้นเท่านั้น !!!

ตัวอย่างตารางความถี่แบบแบ่งเป็นช่วง เช่น

จำนวนชั่วโมงของนักเรียนที่เรียนกับ SMP ใช้ในการอ่านบททบทวนบทเรียนในหนึ่งสัปดาห์

เวลา (ชั่วโมง)	0 – 3	4 – 7	8 – 11	12 – 15	16 – 19
จำนวนคน	30	50	100	120	60

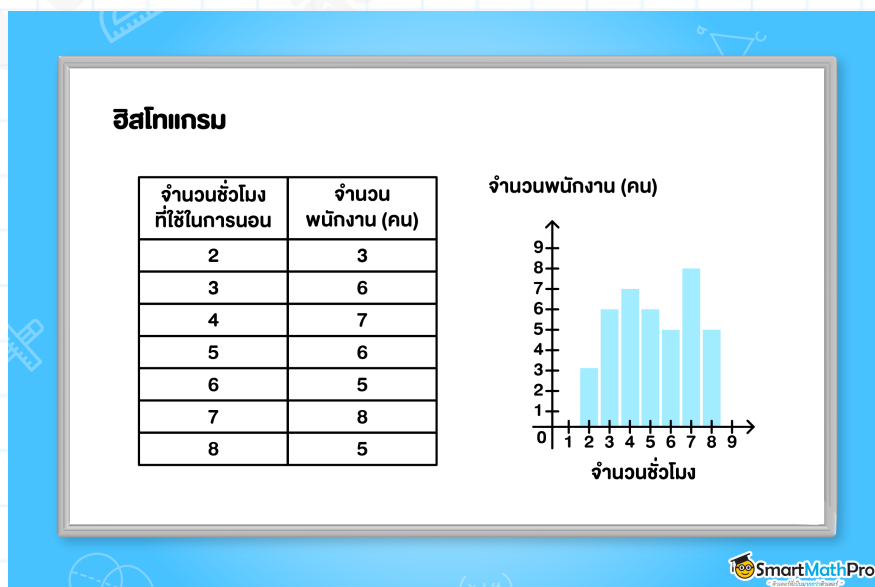
การวิเคราะห์และนำเสนอข้อมูลเชิงปริมาณด้วยแผนภาพ

การใช้ตารางความถี่ในการวิเคราะห์และนำเสนอข้อมูลอาจจะไม่ทำให้เห็นการแจกแจงของข้อมูลมากพอ ดังนั้นเราจึงใช้แผนภาพมาช่วยวิเคราะห์และนำเสนอข้อมูลเชิงปริมาณให้ดีขึ้น มีดังต่อไปนี้

ฮิสโทแกรม (histogram)

เป็นการนำเสนอข้อมูลโดยใช้แท่งสี่เหลี่ยมมุมฉากที่เรียงติดกันบนแกนนอน เมื่อแกนนอนแทนค่าของข้อมูลและความสูงของแท่งแทนความถี่ของข้อมูล คล้าย ๆ กับแผนภูมิแท่ง เช่น

ฝ่ายทรัพยากรบุคคลของบริษัท SMP ได้เก็บข้อมูลจำนวนชั่วโมงที่พนักงานในบริษัทนอนหลับในหนึ่งวันของพนักงานทั้งหมด 40 คนได้ดังนี้



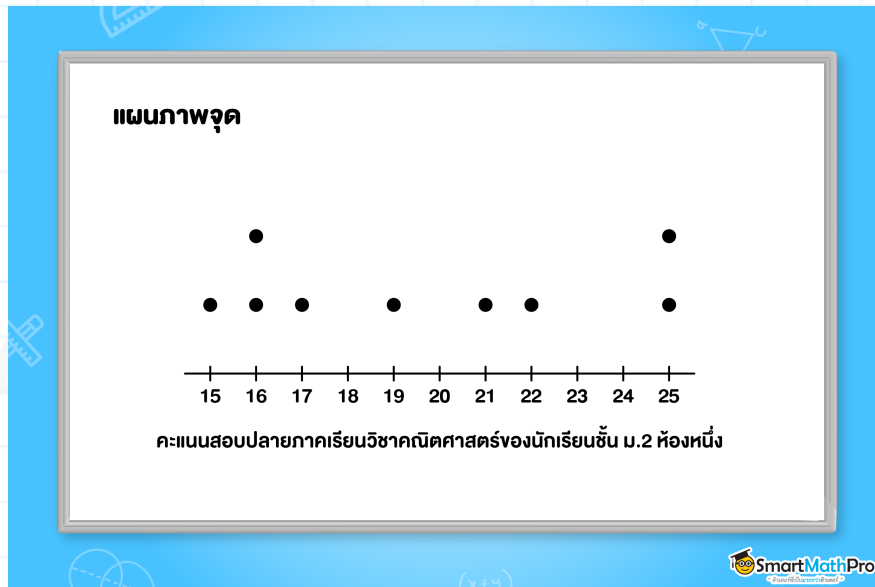
แผนภาพจุด (dot plot)

เป็นการนำเสนอข้อมูลโดยใช้จุดแทนข้อมูลแต่ละตัว เรียงกันในแนวตั้งตรงกับตำแหน่งที่แสดงค่าของข้อมูลแต่ละตัว

เช่น คะแนนสอบปลายภาคเรียนวิชาคณิตศาสตร์ของนักเรียนชั้น ม.2 ห้องหนึ่งเป็นดังนี้

21 25 15 25 16 16 17 22 19

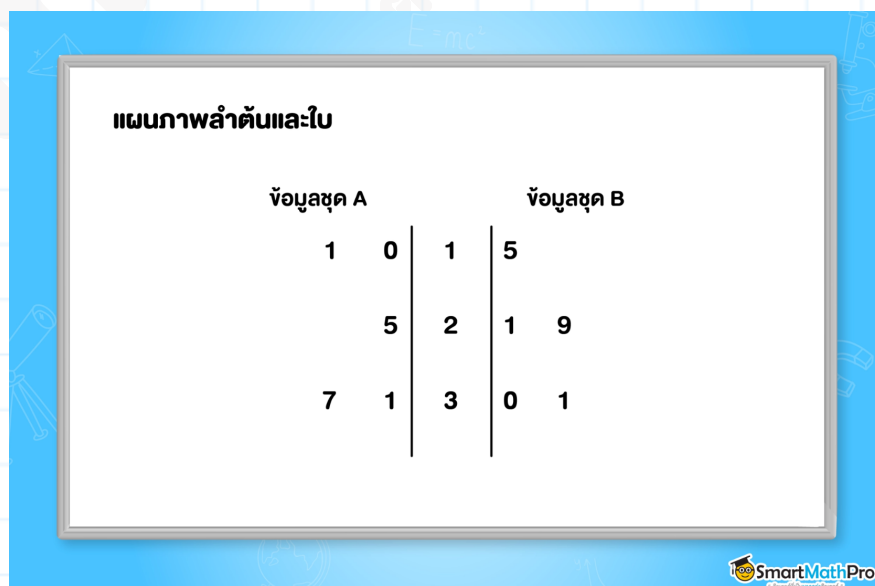
เขียนแผนภาพจุดได้ดังนี้



แผนภาพลำต้นและใบ (stem and leaf plot)

เป็นการนำเสนอข้อมูลโดยใช้แผนภาพ โดยเรียงข้อมูลจากน้อยไปมาก และมีอยู่ด้วยกัน 2 ส่วน คือ ส่วนลำต้น แทนข้อมูลส่วนที่ไม่ใช่หลักหน่วย และส่วนใบ แทนข้อมูลที่เป็นหลักหน่วย เช่น

อายุของคนที่เข้ามาดูข้อมูลชุด A และข้อมูลชุด B บนเว็บไซต์ www.smartmathpro.com เป็นดังนี้



แผนภาพกล่อง (box plot)

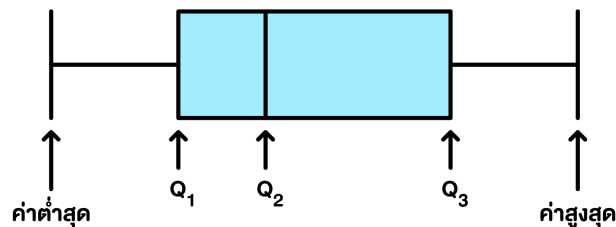
เป็นการนำเสนอข้อมูลที่น้องได้เรียนมาแล้วในสถิติ ม.3 จำได้กันไหมเอ่ย แต่ที่เพิ่มเติมมาจากเนื้อหาในม.ต้น คือ ค่านอกเกณฑ์

ค่านอกเกณฑ์ก็คือค่าที่เกิดขึ้นตามธรรมชาติหรืออาจจะเป็นค่าความคาดเคลื่อนจากการเก็บข้อมูลโดยหาได้จาก

- ค่าที่น้อยกว่า $Q_1 - 1.5(Q_3 - Q_1)$ หรือ
- ค่าที่มากกว่า $Q_1 + 1.5(Q_3 - Q_1)$

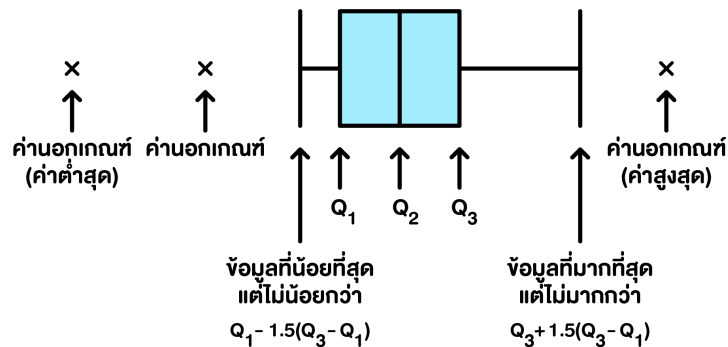
แผนภาพกล่อง

กรณีที่ไม่มีค่านอกเกณฑ์



แผนภาพกล่อง

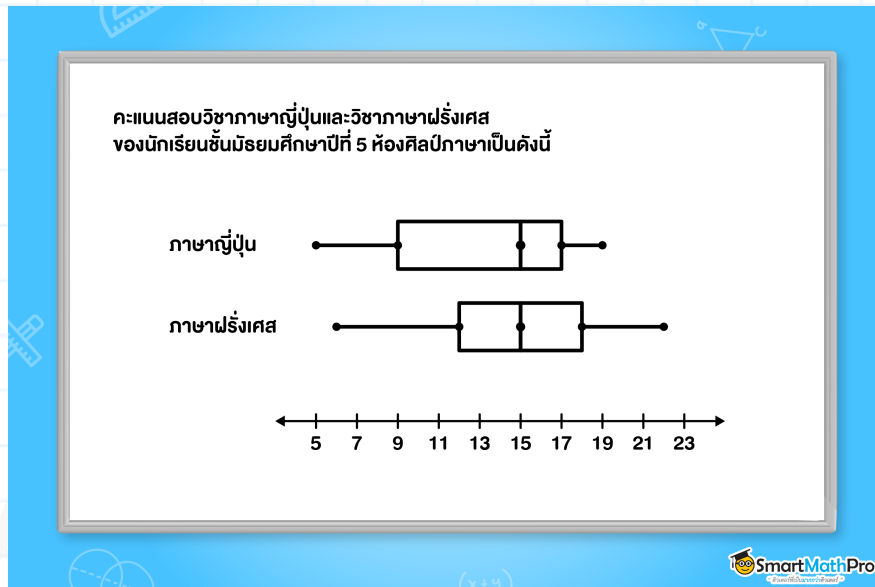
กรณีที่มีค่านอกเกณฑ์



แผนภาพกล่องกับการกระจายข้อมูล

นอกจากที่เราเรียนรู้มาแล้ว ความเจ๋งของแผนภาพกล่องอีกอย่าง คือเราสามารถใช่แผนภาพกล่องในการวิเคราะห์ การกระจายของข้อมูลได้อีกด้วย

ตัวอย่างที่ 2 จากข้อมูลที่กำหนดให้ จงตอบคำถามต่อไปนี้



1. ในวิชาภาษาไทยปุ่น มีนักเรียนชั้น ม.5 ห้องศิลป์ภาษาทั้งหมดกี่เปอร์เซ็นต์ที่ได้คะแนนมากกว่า 15 คะแนน

วิธีทำ

จากแผนภาพจะเห็นว่าที่ 15 คะแนนเป็นตำแหน่ง Q_2 ในแผนภาพกล่อง ของทั้งสองวิชาซึ่ง Q_2 คือตำแหน่งกึ่งกลางของข้อมูลทั้งหมด ดังนั้น จะได้ว่า ในวิชาภาษาไทยปุ่น มีนักเรียนชั้น ม.5 ห้องศิลป์ภาษาทั้งหมด 50% ที่ได้คะแนนมากกว่า 15 คะแนน

2. จากแผนภาพ คะแนนช่วงไหนของวิชาภาษาไทยปุ่นและวิชาภาษาอังกฤษมีการกระจายตัวมากที่สุด

วิธีทำ

ข้อมูลในแต่ละช่วงของแผนภาพกล่องจะมีจำนวน 25% ของจำนวนข้อมูลทั้งหมด ดังนั้นในการพิจารณาการกระจายของข้อมูล เราจะดูจากความกว้างของแต่ละช่วง หากช่วงไหนมีความกว้างมากก็จะได้ว่ามีการกระจายตัวมาก

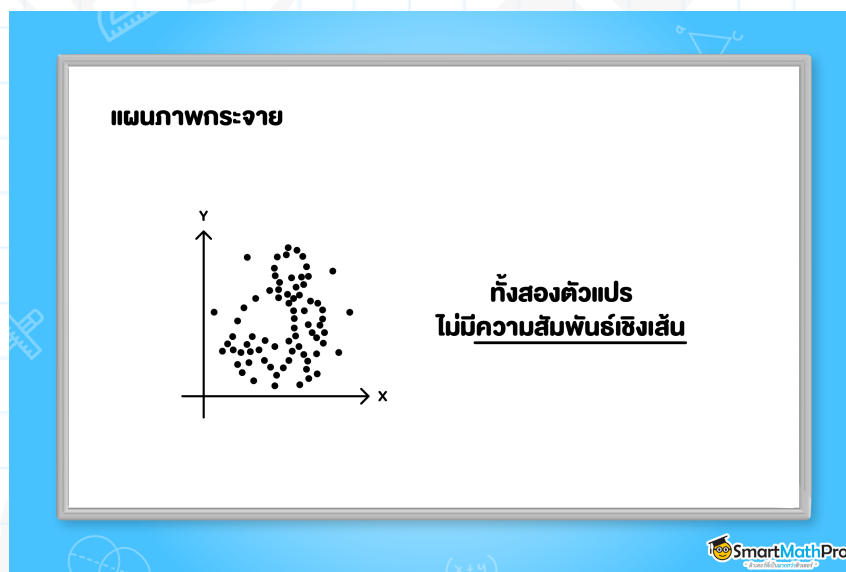
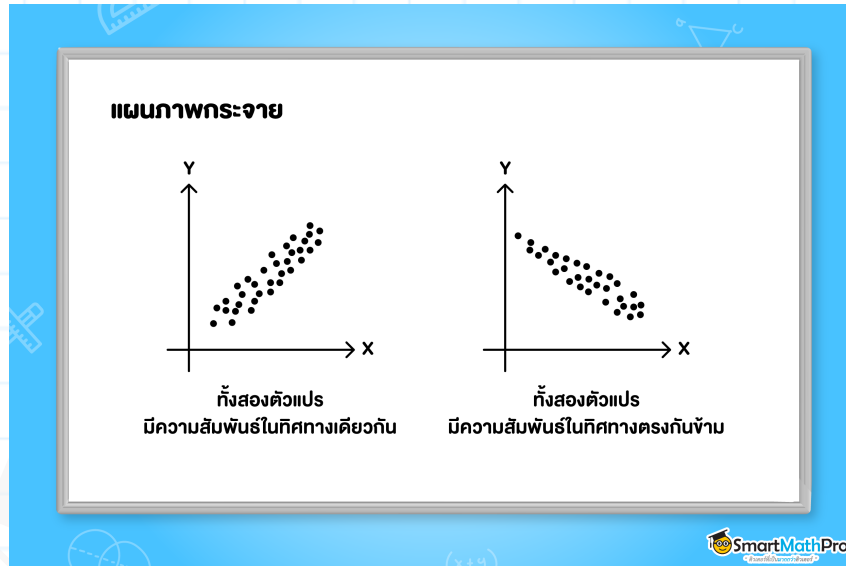
พิจารณาวิชาภาษาไทยปุ่น จะเห็นว่าช่วง Q_1 ถึง Q_2 มีความกว้างของช่วงมากที่สุด ดังนั้น ในวิชาภาษาไทยปุ่น คะแนนระหว่าง 9 ถึง 15 มีการกระจายตัวมากที่สุด

พิจารณาวิชาภาษาอังกฤษ จะเห็นว่าช่วง ค่าต่ำสุด ถึง Q_1 มีความกว้างของช่วงมากที่สุด ดังนั้น ในวิชาภาษาอังกฤษ คะแนนระหว่าง 6 ถึง 12 มีการกระจายตัวมากที่สุด

จากแผนภาพที่ผ่านมาเป็นการนำเสนอข้อมูลที่มีตัวแปรเพียงหนึ่งตัวเท่านั้น ถ้าหากเรามีข้อมูลที่ไม่ได้มีเพียงตัวแปรเดียว หรือไม่รู้ว่ตัวแปรไหนเป็นตัวแปรต้นหรือตัวแปรตาม เราก็จะใช้แผนภาพต่อไปนี้

แผนภาพการกระจาย (scatter plot)

คือ การนำเสนอข้อมูลที่มีสองตัวแปรในรูปแบบการกระจายของจุด ซึ่งจะแสดงให้เห็นถึงความสัมพันธ์ของสองตัวแปรนั้น



ค่าวัดทางสถิติ

ค่ากลางของข้อมูล

น้อง ๆ ได้ศึกษาเรื่องค่าเฉลี่ยเลขคณิต มัธยฐาน และฐานนิยมกันมาแล้วในสถิติระดับ ม.ต้น ซึ่งค่าทั้งสามนี้เรียกว่า ค่ากลางของข้อมูล แต่ละค่าก็จะมีข้อดี ข้อเสีย และความเหมาะสมในการนำไปใช้ที่แตกต่างกัน

ค่าเฉลี่ยเลขคณิต

ค่าเฉลี่ยเลขคณิต (arithmetic mean) เป็นค่าที่หาได้จากการหารผลรวมของข้อมูลทั้งหมด ด้วยจำนวนข้อมูลที่มี

สูตรในการหาค่าเฉลี่ยเลขคณิตจะมี 2 สูตร คือสูตรที่ใช้สำหรับประชากรและตัวอย่าง
 นอง ๆ อย่าลืมอ่านโจทย์ให้ดีก่อนเลือกใช้สูตร

ค่ากลางของข้อมูล

ค่าเฉลี่ยเลขคณิตของประชากร (μ)	ค่าเฉลี่ยเลขคณิตของตัวอย่าง ($\bar{x}$)
$\mu = \frac{\sum_{i=1}^N x_i}{N} = \frac{x_1 + x_2 + x_3 + \dots + x_N}{N}$	$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n}$
เมื่อ $x_1, x_2, x_3, \dots, x_N$ แทนข้อมูล N แทนขนาดประชากร	เมื่อ $x_1, x_2, x_3, \dots, x_n$ แทนข้อมูล n แทนขนาดตัวอย่าง

ค่าเฉลี่ยเลขคณิตถ่วงน้ำหนัก

ค่าเฉลี่ยเลขคณิตถ่วงน้ำหนัก (weighted arithmetic mean) เหมาะสำหรับการใช้กรณีที่
 แต่ละค่ามีความสำคัญไม่เท่ากัน

เช่น การคำนวณเกรดเฉลี่ย (GPA) ซึ่งแต่ละวิชามีหน่วยกิตไม่เท่ากัน โดยหาได้จากสูตร

$$\text{ค่าเฉลี่ยเลขคณิตถ่วงน้ำหนัก} = \frac{w_1x_1 + w_2x_2 + w_3x_3 + \dots + w_Nx_N}{w_1 + w_2 + w_3 + \dots + w_N}$$

เมื่อ $x_1, x_2, x_3, \dots, x_N$ แทนข้อมูล

N แทนขนาดประชากร และ $w_1, w_2, w_3, \dots, w_N$ แทนน้ำหนักของข้อมูล $x_1, x_2, x_3, \dots, x_N$ ตามลำดับ

ตัวอย่างที่ 3

สมมติหะเบียนเรียน 4 วิชา ซึ่งแต่ละวิชามีหน่วยกิตไม่เท่ากัน และสมมติได้เกรดแต่ละวิชาดังนี้

วิชา	คณิต	ไทย	สังคม	อังกฤษ
เกรด	4.0	4.0	2.5	3.5
หน่วยกิต	2	1	1	1

จงหาเกรดเฉลี่ยของสม

วิธีทำ

$$\text{จาก ค่าเฉลี่ยเลขคณิตถ่วงน้ำหนัก} = \frac{w_1x_1 + w_2x_2 + w_3x_3 + \dots + w_Nx_N}{w_1 + w_2 + w_3 + \dots + w_N}$$

$$\begin{aligned}\text{จะได้ เกรดเฉลี่ยของส้ม} &= \frac{(2)(4.0) + (1)(4.0) + (1)(2.5) + (1)(3.5)}{2 + 1 + 1 + 1} \\ &= \frac{18}{5} \\ &= 3.60\end{aligned}$$

มัธยฐาน

มัธยฐาน คือ ข้อมูลที่อยู่ในตำแหน่งกึ่งกลาง แต่ระวังว่าก่อนนั้นเอง ๆ จะหาว่าข้อมูลไหนอยู่ตำแหน่งกึ่งกลาง จะต้องเรียงข้อมูลจากน้อยไปมาก หรือมากไปน้อยด้วย

ถ้าข้อมูลมีจำนวนมาก เราคงไม่สะดวกใช้วิธีการมองหาข้อมูลในตำแหน่งกึ่งกลางตรง ๆ
น้องสามารถใช้สูตรนี้ได้

ถ้าข้อมูลมี n ตัว และเรียงข้อมูลจากน้อยไปมากหรือมากไปน้อยแล้ว
จะได้ว่ามัธยฐานอยู่ในตำแหน่ง $\frac{n+1}{2}$

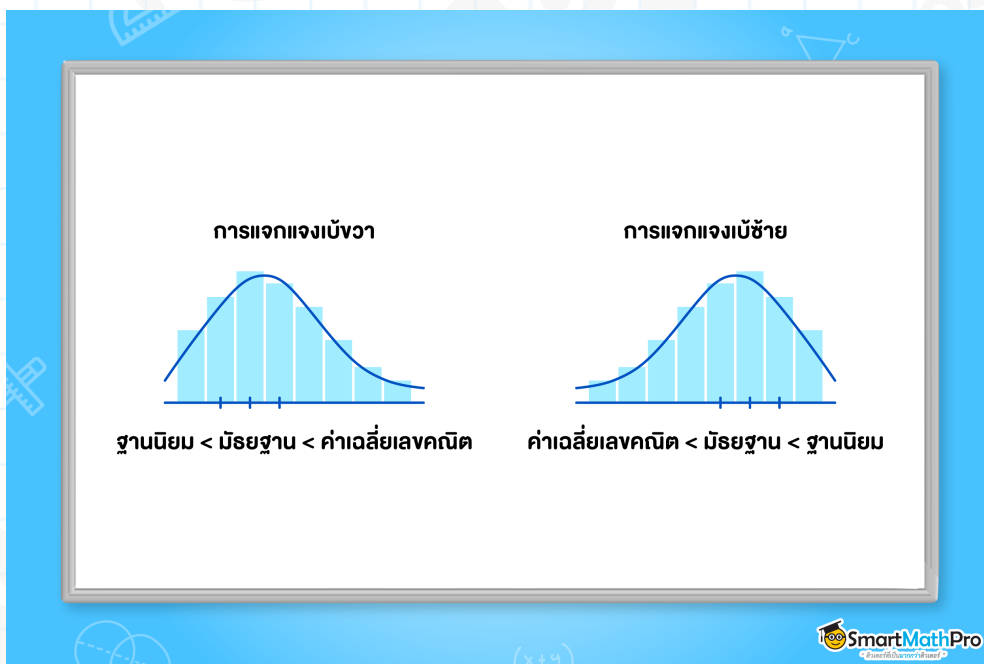
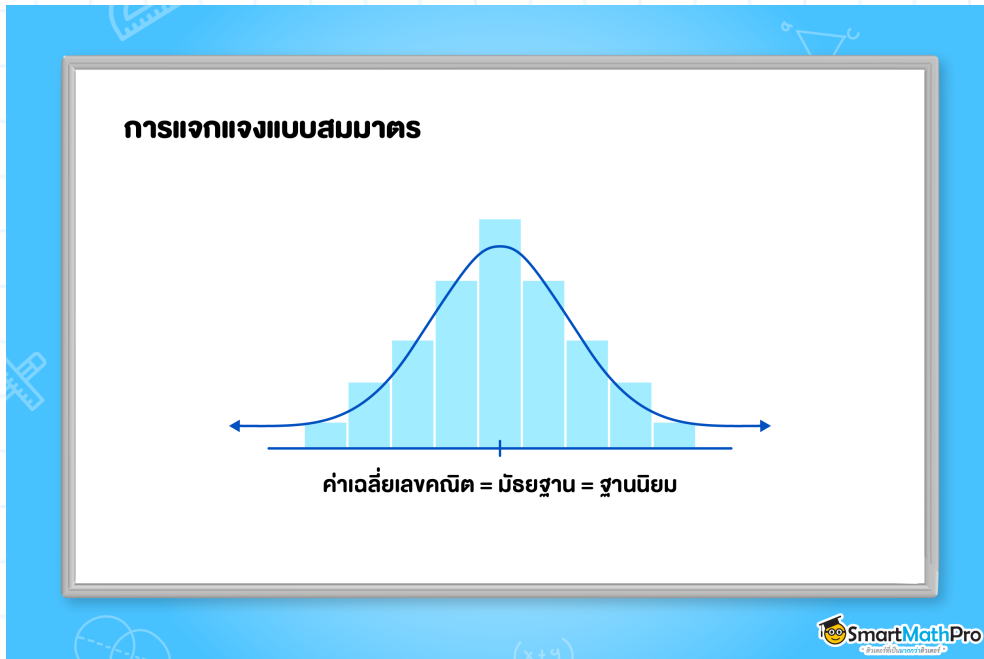
ระวังนิดนึงว่า น้อง ๆ บางคนแทนค่า n หา $\frac{n+1}{2}$ แล้วเอาไปตอบเลยจะผิด เพราะค่าที่ได้เป็นตำแหน่งของมัธยฐาน ไม่ใช่ค่าของข้อมูลที่เป็นมัธยฐาน

ฐานนิยม

ฐานนิยม คือ ข้อมูลที่มีจำนวนครั้งของการเกิดซ้ำมากที่สุด ซึ่งน้อง ๆ ที่อ่านมาถึงตรงนี้อาจจะคิดว่าเราสามารถหาฐานนิยมในข้อมูลเชิงคุณภาพได้ และถูกต้องแล้วนะ เราสามารถหาฐานนิยมในข้อมูลเชิงปริมาณได้เหมือนกัน

การกระจายของข้อมูล

รูปร่างความสัมพันธ์ระหว่างการกระจายข้อมูลและค่ากลางของข้อมูลแบ่งได้เป็น 3 แบบ คือ การแจกแจงแบบสมมาตร การแจกแจงเบ้ขวา และการแจกแจงเบ้ซ้าย ตามภาพด้านล่างนี้เลยนะ



- การแจกแจงแบบสมมาตร : โค้งโด่งตรงกลาง แปลว่าข้อมูลหนาแน่นตรงกลาง
- การแจกแจงเบ้ขวา : โค้งโด่งทางซ้าย ข้อมูลทางซ้ายหนาแน่น แปลว่าข้อมูลเบ้ขวา
- การแจกแจงเบ้ซ้าย : โค้งโด่งทางขวา ข้อมูลทางขวาหนาแน่น แปลว่าข้อมูลเบ้ซ้าย

ค่าวัดการกระจาย

โดยทั่วไป การวัดการกระจายแบ่งได้ 2 วิธี คือ การกระจายสัมบูรณ์ (absolute variation) และการกระจายสัมพัทธ์ (relative variation)

การกระจายสัมบูรณ์ ในระดับชั้นนี้จะศึกษาการวัดการกระจายสัมบูรณ์ 4 ชนิด คือ

1. พิสัย

พิสัย (range) คือ ค่าที่ใช้วัดการกระจายของข้อมูลชุดหนึ่ง โดยหาได้จากผลต่างระหว่างค่าสูงสุดและค่าต่ำสุดของข้อมูลชุดนั้น

กำหนดให้ข้อมูลหนึ่งมี x_{max} และ x_{min} เป็นค่าสูงสุดและค่าต่ำสุด ตามลำดับ

$$\text{พิสัย} = x_{max} - x_{min}$$

2. พิสัยระหว่างควอร์ไทล์

พิสัยระหว่างควอร์ไทล์ (interquartile range) คือ ค่าที่ใช้วัดการกระจายของข้อมูลโดยใช้ผลต่างระหว่างควอร์ไทล์ที่สาม และควอร์ไทล์ที่หนึ่ง เขียนแทนพิสัยระหว่างควอร์ไทล์ด้วย IQR

ให้ Q_1 และ Q_3 เป็นควอร์ไทล์ที่หนึ่งและควอร์ไทล์ที่สาม ของข้อมูลชุดหนึ่ง จะได้

$$IQR = Q_3 - Q_1$$

น้อง ๆ ที่อ่านมาถึงตรงนี้น่าจะเคยเห็นสัญลักษณ์ IQR มาแล้วจากการหาด่านอกเกณฑ์ในเรื่องแผนภาพกล่อง ใช่แล้วว IQR ที่หมายถึงพิสัยระหว่างควอร์ไทล์ในเรื่องนี้ กับ IQR ในสูตรการหาด่านอกเกณฑ์เป็นอันเดียวกันเลย

3. ส่วนเบี่ยงเบนมาตรฐาน

ส่วนเบี่ยงเบนมาตรฐาน (standard deviation) คือ ค่าที่ใช้ในการวัดการกระจายของข้อมูล โดยที่บอกให้เราทราบว่าข้อมูลแต่ละตัวอยู่ห่างจากค่าเฉลี่ยเลขคณิตอยู่ประมาณเท่าไร

ส่วนเบี่ยงเบนมาตรฐาน

ส่วนเบี่ยงเบนมาตรฐานของประชากร (σ)	ส่วนเบี่ยงเบนมาตรฐานของตัวอย่าง (s)
$\sigma = \sqrt{\frac{\sum_{i=1}^N (x_i - \mu)^2}{N}}$	$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$
เมื่อ $x_1, x_2, x_3, \dots, x_N$ แทนข้อมูล N แทนขนาดประชากร และ μ แทนค่าเฉลี่ยเลขคณิต	เมื่อ $x_1, x_2, x_3, \dots, x_n$ แทนข้อมูล n แทนขนาดตัวอย่าง และ $\bar{x}$ แทนค่าเฉลี่ยเลขคณิต

4. ความแปรปรวน

ความแปรปรวน (variance) คือ ค่าที่ใช้วัดการกระจายของข้อมูล โดยหากกำลังสองของ ส่วนเบี่ยงเบนมาตรฐาน จะได้สูตรการหาความแปรปรวนเป็นดังนี้

ความแปรปรวน	
ความแปรปรวนของประชากร (σ^2)	ความแปรปรวนของตัวอย่าง (s^2)
$\sigma^2 = \frac{\sum_{i=1}^N (x_i - \mu)^2}{N}$	$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$
เมื่อ $x_1, x_2, x_3, \dots, x_N$ แทนข้อมูล N แทนขนาดประชากร และ μ แทนค่าเฉลี่ยเลขคณิต	เมื่อ $x_1, x_2, x_3, \dots, x_n$ แทนข้อมูล n แทนขนาดตัวอย่าง และ $\bar{x}$ แทนค่าเฉลี่ยเลขคณิต

การกระจายสัมพัทธ์

หากน้อง ๆ มีข้อมูล 2 ชุด เช่น ถ้าข้อมูลของเรามีค่าน้อยสัก 0 – 10 สมมติว่าคำนวณค่าของ ส่วนเบี่ยงเบนมาตรฐานได้ 2 และข้อมูลอีกชุดหนึ่งมีค่าอยู่ในช่วง 500 – 1,000 มีส่วนเบี่ยงเบน มาตรฐานเป็น 80 เราไม่สามารถสรุปได้ว่าการกระจายของข้อมูลชุดที่สองนั้นมากกว่าชุดแรก เพราะค่าของข้อมูลต่างกันมาก

พี่จะแนะนำให้น้อง ๆ ใช้สัมประสิทธิ์การแปรผัน (coefficient of variation) มาเปรียบเทียบ การกระจายของข้อมูลแต่ละชุดมากกว่าการใช้ส่วนเบี่ยงเบนมาตรฐานหรือการกระจายสัมบูรณ์ โดยหาได้จากสูตรต่อไปนี้

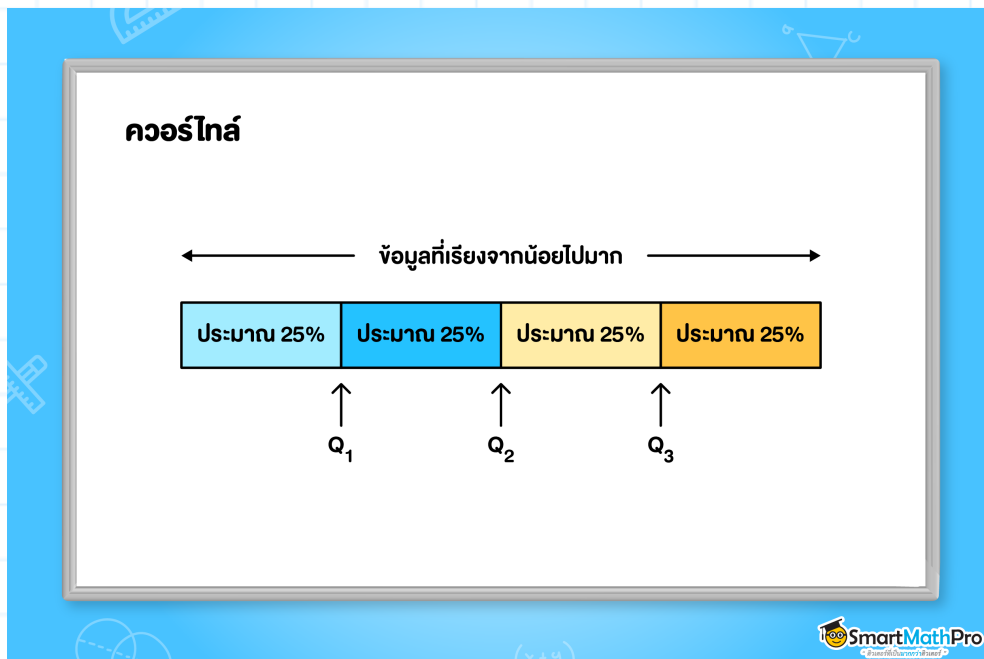
สัมประสิทธิ์การแปรผันของประชากร	สัมประสิทธิ์การแปรผันของตัวอย่าง
$\frac{\sigma}{ \mu }$ เมื่อ $\mu \neq 0$	$\frac{s}{ \bar{x} }$ เมื่อ $\bar{x} \neq 0$

การวัดตำแหน่งของข้อมูล

น้อง ๆ มีพื้นฐานการวัดตำแหน่งของข้อมูลกันมาแล้ว นั่นก็คือ ควอร์ไทล์ แต่จะมีการวัดตำแหน่ง ของข้อมูลอีกอันหนึ่งที่เพิ่มขึ้นมา นั่นก็คือ เปอร์เซ็นไทล์

ควอร์ไทล์

ควอร์ไทล์ (Quartile) จะแบ่งข้อมูลออกเป็น 4 ส่วนเท่า ๆ กัน ให้น้องเรียงข้อมูลจากน้อยไปมาก ซึ่งหากพิจารณาข้อมูลทั้งหมดเป็น 100% แต่ละส่วนนั้นจะมีจำนวนข้อมูลประมาณ 25% ของข้อมูลทั้งหมด



พี่แนะนำให้ทุกคนหาควอร์ไทล์โดยใช้สูตร ซึ่งสะดวกกว่าการใช้วิธีการหามัธยฐานไปเรื่อย ๆ ในสมัยที่น้องเรียน ม.3 น้า

วิธีการหาควอร์ไทล์

1. เรียงข้อมูลจากน้อยไปมาก
2. หาดำแหน่งของควอร์ไทล์ที่ i โดยตำแหน่งของ Q_i คือ $\frac{i(n+1)}{4}$ เมื่อ n คือจำนวนของข้อมูล
3. นำตำแหน่งไปพิจารณาหาค่าของ Q_i

เปอร์เซ็นต์ไทล์

การหา เปอร์เซ็นต์ไทล์ (Percentile) ทำได้ในทำนองเดียวกันกับการหาควอร์ไทล์ แตกต่างกันเพียงสูตรการหาดำแหน่ง โดยตำแหน่งของเปอร์เซ็นต์ไทล์ที่ i หรือ ตำแหน่งของ P_i คือ $\frac{i(n+1)}{100}$

“เวลาที่ดียิ่งที่สุดในการเริ่มต้น คือ ตอนนี่”

- พี่ป๊าน SmartMathPro -

สนใจติวคณิตศาสตร์เพิ่มเติม online.smartmathpro.com