1. แนวคิดพื้นฐานทางสถิติ

The Basic Idea of Statistics

คำว่า "สถิติ" มีความหมาย 2 ประการ คือ

1. ตัวเลขสถิติ (Statistics) หมายถึงจำนวนหรือค่าที่ได้จากการรวบรวมข้อมูล แสดงถึงข้อเท็จจริงของสิ่งต่างๆ อย่างมีความหมาย เช่น ค่าเฉลี่ยของคะแนนวิชาคณิตศาสตร์ของนักเรียน ม.5 ภาคเรียนที่ 2 ปีการศึกษา 2555

2. วิชาสถิติหรือสถิติศาสตร์ (Statistics) หมายถึงศาสตร์ว่าด้วยระเบียบวิธีทางสถิติ ซึ่งประกอบด้วย 4 ขั้นตอน คือ **การเก็บรวบรวมข้อมูล การนำเสนอข้อมูล การวิเคราะห์ข้อมูล และการแปลความหมายข้อมูล**

ประเภทของสถิติ

ในวิชาสถิติ แบ่งสถิติออกเป็น 2 ประเภทใหญ่ๆ คือ สถิติพรรณนา และสถิติอนุมาน

1. สถิติพรรณนา (Descriptive statistics) หรือเรียกอีกอย่างหนึ่งว่า "สถิติบรรยาย" เป็นสถิติที่มุ่งศึกษาเพื่ออธิบายเรื่องราวต่างๆ ของกลุ่มประชากร (Population) กลุ่มใดกลุ่มหนึ่งโดยเฉพาะอาจเป็นกลุ่มใหญ่หรือกลุ่มเล็กก็ได้ โดยทำการเก็บรวบรวมข้อมูลจากสมาชิกทุกหน่วยในกลุ่มประชากรนั้น ผลการศึกษาใช้อธิบายหรือสรุปเกี่ยวกับเรื่องราวของกลุ่มที่ศึกษาเท่านั้น ไม่สามารถนำผลการศึกษาไปสรุปอ้างอิงถึงกลุ่มอื่นๆ ที่ไม่ได้ศึกษา

ตัวอย่าง เช่น ฝ่ายแนะแนวของโรงเรียนมหิดลวิทยานุสรณ์ การวัดสติปัญญา (I.Q.) ของนักเรียนชั้นมัธยมศึกษาปีที่ 5 ทุกคนของโรงเรียนนี้ ได้ค่าเฉลี่ยเท่ากับ 115 เป็นค่าเฉลี่ยของนักเรียนชั้นมัธยมศึกษาปีที่ 5 ของโรงเรียนนี้เท่านั้น จะสรุปว่านักเรียนชั้นมัธยมศึกษาปีที่ 5 ของโรงเรียนอื่นหรือของจังหวัด มีระดับสติปัญญาเฉลี่ยเท่ากับ 115 ด้วยไม่ได้

สถิติพรรณนา มีหลายชนิด ได้แก่

1.1 สถิติพื้นฐาน เช่น ความถี่ สัดส่วน ร้อยละ

1.2 การวัดตำแหน่ง เช่น อันดับที่ ควอไทล์ เดไซล์ เปอร์เซ็นไทล์

1.3 การวัดแนวโน้มเข้าสู่ส่วนกลาง เช่น ฐานนิยม มัธยฐาน ค่าเฉลี่ยเลขคณิต ค่าเฉลี่ยเรขาคณิต ค่าเฉลี่ยฮาร์โมนิค

1.4 การวัดการกระจาย เช่น พิสัย พิสัยควอไทล์ ส่วนเบี่ยงเบนเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน ความแปรปรวน สัมประสิทธิ์การกระจาย

1.5 การวัดความสัมพันธ์ เช่น สัมประสิทธิ์ถดถอย

ค่าต่างๆ ที่คำนวณได้จากข้อมูล ที่เก็บรวบรวมจากสมาชิกทุกๆ หน่วยของกลุ่มประชากร เรียกว่า **ค่าแท้** หรือค่า **พารามิเตอร์** (Parameter) มีคุณสมบัติเป็นค่าคงที่ (Constant)

2. สถิติอ้างอิง (Inferential statistics) หรือที่เรียกอีกอย่างหนึ่งว่า **สถิติอนุมาน** เป็นสถิติที่มุ่งศึกษาเพื่อหาข้อสรุปเรื่องราวของประชากร โดยเก็บรวบรวมข้อมูลจากกลุ่มย่อยที่เรียกว่า **กลุ่มตัวอย่าง** (Sample) แล้วนำผลการศึกษาไปสรุปอ้างอิงถึงกลุ่มใหญ่ที่เรียกว่า **กลุ่มประชากร** (Population) ซึ่งเป็นกลุ่ม เป้าหมายที่ต้องการศึกษา

ค่าต่างๆ ที่คำนวณได้จากข้อมูลที่เก็บรวบรวมจากกลุ่มตัวอย่าง เรียกว่า **ค่าสถิติ** (Statistic) มีคุณสมบัติเป็นตัวแปร (Variable)

สถิติอ้างอิงแบ่งออกเป็น 2 ประเภท คือ การประมาณค่าและการทดสอบสมมติฐาน

2.1 การประมาณค่า (Estimation) เป็นการประมาณค่าแท้ของประชากร เรียกว่าค่าพารามิเตอร์ (Parameter) โดยใช้ค่าสถิติ (Statistic) ที่ได้จากกลุ่มตัวอย่าง

2.2 การทดสอบสมมติฐาน (Testing Statistical Hypothesis) เป็นการทดสอบสมมติฐานทางสถิติ เพื่อสรุปอ้างอิงค่าสถิติต่างๆ ไปยังกลุ่มประชากร กล่าวได้อีกอย่างหนึ่งว่า เป็นการทดสอบความมีนัยสำคัญทางสถิติ

ความหมายของคำที่เกี่ยวข้องกับวิชาสถิติ

ประชากร (Population) หมายถึงกลุ่มของสมาชิกทุกหน่วยที่เราต้องการศึกษาลักษณะ (Characteristics) บางอย่าง ตัวอย่างเช่น ถ้าต้องการศึกษาอายุการใช้งานของหลอดไฟในโรงงานแห่งหนึ่ง ประชากร คือ จำนวนหลอดไฟทั้งหมดในโรงงานนี้นั่นเอง

พารามิเตอร์ (Parameter) หมายถึง ตัวเลขซึ่งแสดงคุณสมบัติบางประการของประชากร ตัวอย่างเช่น ค่าเฉลี่ยของประชากร (μ) ความแปรปรวนของประชากร (σ^2) ค่าส่วนเบี่ยงเบนมาตรฐานของประชากร (σ) เป็นต้น

ตัวอย่าง (Sample) หมายถึง กลุ่มย่อยของสมาชิกในกลุ่มประชากรที่เลือกมาเพื่อศึกษาลักษณะที่สนใจ เช่น ศึกษาอายุการใช้งานหลอดไฟ อาจจะสุ่มตัวอย่างหลอดไฟฟ้ามา 20 หลอด จากทั้งหมด 1000 หลอดในโรงงานนั้น เป็นต้น

ค่าสถิติ (Statistic) หมายถึง การวัดผลที่ได้จากตัวอย่าง ซึ่งพรรณนาลักษณะของตัวอย่าง เช่น ค่าเฉลี่ยของตัวอย่าง ($\bar{x}$) ค่าความแปรปรวนของตัวอย่าง (s^2) ค่าส่วนเบี่ยงเบนมาตรฐานของตัวอย่าง (s) เป็นต้น

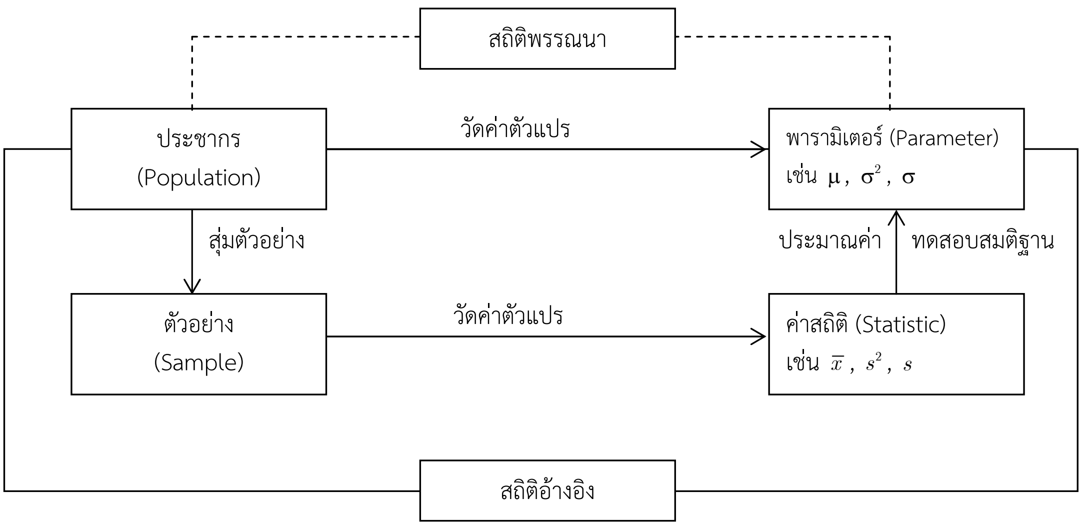
ข้อมูล (Data) หมายถึง ข้อเท็จจริงที่เป็นตัวเลขหรือข้อความที่รวบรวมมาได้จากเรื่องที่เรากำลังศึกษา

ตัวแปร (Variable) หมายถึง ลักษณะของประชากรที่เราสนใจวิเคราะห์ โดยลักษณะนั้นๆ สามารถเปลี่ยนค่าได้ เช่น เพศ คะแนนสอบ

ประโยชน์ของสถิติ

1. เป็นสิ่งชี้ให้เห็นถึงข้อเท็จจริงของเหตุการณ์ และเรื่องราวที่สนใจอยู่
2. เป็นเครื่องมือในการวางแผนงานของโครงการและกิจการต่างๆ
3. เป็นระเบียบวิธีสำหรับการวิเคราะห์ในงานวิจัยโดยทั่วๆ ไป
4. เป็นเครื่องมือในการประเมินผลงานที่ได้ทำไปแล้ว

แผนภูมิแสดงมโนทัศน์เกี่ยวกับสถิติพรรณนาและสถิติอ้างอิง



2. การเก็บรวบรวมข้อมูล

Collecting Data Sensibly

ข้อมูล หมายถึง ข้อความจริงซึ่งอาจเป็นตัวเลขหรืออาจไม่ใช่ตัวเลขก็ได้ การทำความเข้าใจข้อมูลที่เก็บรวบรวมมาได้นั้นมีความสำคัญมาก เพราะจะนำไปสู่รูปแบบการวิเคราะห์ข้อมูล ดังนั้นหากพิจารณาถึงข้อดีข้อด้อยของข้อมูลทั้งสองประเภท กล่าวคือ **ข้อมูลเชิงปริมาณ** เป็นข้อมูลที่บอกลักษณะด้วยตัวเลขได้ดี มีข้อดี คือ เข้าใจง่าย สะดวก เนื่องจากเราสามารถตีความได้ตามตัวเลข ส่วนข้อด้อย คือ อาจเกิดความคลาดเคลื่อนได้จากตัวเลข จากวิธีเก็บรวบรวม จากลักษณะของตัวเลขเองที่มีรูปแบบและขนาด รวมถึงหน่วยที่แตกต่างกัน **ข้อมูลเชิงคุณภาพ** เป็นข้อมูลที่บอกลักษณะทางธรรมชาติของสังคมซึ่งมีความเกี่ยวโยงกันมากมาย ข้อดี คือ สามารถอธิบาย สะท้อนความคิด ความรู้สึกมีรายละเอียดมากผสมผสานกันเป็นส่วนเดียวกัน หรืออาจจะแยกเป็นประเด็นต่างๆ ขึ้นอยู่กับเป้าหมายที่ต้องการ สามารถใช้ข้อมูลที่มีอยู่ในสภาพปกติ เช่น ระเบียบ ภาพถ่าย เทปบันทึกเสียง เทปบันทึกภาพ และบันทึกเอกสารต่างๆ ประกอบได้ ส่วนข้อด้อย ถ้าผู้สนใจศึกษามีอคติจะขาดความตรงและความเชื่อมั่น ไม่มีแบบตายตัว ขึ้นอยู่กับความสามารถของผู้วิจัย อาจจะใช้เวลาในการเก็บรวบรวมข้อมูลนานและวิเคราะห์ข้อมูลได้ยาก

2.1 ประเภทของข้อมูล

จำแนกประเภทตามวิธีการเก็บรวบรวมแบ่งเป็น 2 ประเภท

1. ข้อมูลปฐมภูมิ คือข้อมูลที่ผู้ใช้จะต้องเก็บรวบรวมจากผู้ให้ข้อมูล หรือแหล่งที่มาของข้อมูลโดยตรง
2. ข้อมูลทุติยภูมิ คือข้อมูลที่ผู้ใช้ไม่ต้องเก็บรวบรวมจากผู้ให้ข้อมูลหรือแหล่งที่มาของข้อมูลโดยตรง แต่ได้มาจากข้อมูลที่ผู้อื่นเก็บรวบรวมไว้แล้ว

จำแนกตามลักษณะของข้อมูลมี 2 ลักษณะ

1. ข้อมูลเชิงปริมาณ คือข้อมูลที่ใช้แทนขนาดหรือปริมาณซึ่ง วัดออกมาเป็นค่าตัวเลขที่สามารถนำมาใช้เปรียบเทียบขนาดได้โดยตรง
2. ข้อมูลเชิงคุณภาพ คือข้อมูลที่ไม่สามารถวัดออกมาเป็นค่า ตัวเลขได้โดยตรง แต่วัดออกมาในเชิงคุณภาพได้

2.2 วิธีเก็บรวบรวมข้อมูล

การเก็บรวบรวมข้อมูลปฐมภูมิ มี 2 วิธีคือ

การสำมะโน ซึ่งอาจทำได้โดย

- การสัมภาษณ์
- การสอบถามทางไปรษณีย์
- การสอบถามทางโทรศัพท์
- การสังเกต
- การทดลอง

การสำรวจจากกลุ่มตัวอย่าง การเลือกตัวอย่างจากประชากรอาจทำได้โดย

- เลือกตัวอย่างชนิดที่ไม่ทราบโอกาส หรือความน่าจะเป็นที่แต่ละหน่วยถูกเลือกขึ้นมาเป็นตัวอย่าง
- เลือกตัวอย่างชนิดที่ทราบโอกาส หรือความน่าจะเป็นที่แต่ละหน่วยถูกเลือกขึ้นมาเป็นตัวอย่าง

การเก็บรวบรวมข้อมูลทุติยภูมิ ควรดำเนินการดังนี้

- พิจารณาตัวบุคคลผู้เขียนรายงาน บทความว่า มีความรู้เชี่ยวชาญเรื่องใด
- ควรรวบรวมข้อมูลจากหลายๆ แหล่ง เพื่อเปรียบเทียบและกันความผิดพลาด
- พิจารณาจากลักษณะของข้อมูลที่ต้องการเก็บรวบรวมว่าเป็นข้อมูลที่เป็นความจริง หรือความเห็น
- ถ้าข้อมูลที่จะเก็บรวบรวมได้มาจากการสำรวจตัวอย่างหรือต้องผ่านการวิเคราะห์มาแล้ว ควรพิจารณาถึงวิธีการที่ใช้ว่าเหมาะสมหรือไม่

แบบฝึกทักษะ 2

1. ข้อมูลต่อไปนี้เป็นข้อมูลเชิงปริมาณหรือข้อมูลเชิงคุณภาพ
    a. คะแนนสอบวิชาคณิตศาสตร์ --
    b. จำนวนผู้โดยสารที่รอรถ --
    c. เลขทะเบียนรถ --
    d. หมายเลขโทรศัพท์ --
    e. ราคาข้าวสารต่อกิโลกรัม --
    f. หมายเลขประจำตัวนักเรียน --
    g. ขนาดรองเท้าของนักเรียน --
    h. รายได้ของคนในครอบครัว --

2. "สำนักวิจัยโรงเรียนมหิดลวิทยานุสรณ์เปิดเผยว่า ผลการสำรวจความคิดเห็นของประชากรในจังหวัดนครปฐมต่อการยอมรับมาตรการประหยัดน้ำมัน พบว่าส่วนใหญ่ยอมรับที่จะปฏิบัติตามมาตรการเหล่านั้น" จากข้อความข้างต้น
    a. กลุ่มประชากรคือ --
    b. กลุ่มตัวอย่างคือ --

3. จงพิจารณาว่าควรใช้วิธีในการเก็บข้อมูลต่อไปนี้
    a. รายได้เฉลี่ยต่อครัวเรือนของคนในจังหวัดนครปฐม --
    b. จำนวนผู้ใช้บริการห้องสมุดโรงเรียนมหิดลวิทยานุสรณ์ --
    c. ผลการเคลือบฟลูออไรด์บนผิวฟันที่มีผลต่อการป้องกันฟันผุ --
    d. การเปรียบเทียบประสิทธิภาพของยาแก้ปวด --
    e. การสำรวจความคิดเห็นกรณีคลื่นยักษ์ถล่มภาคใต้ --

3. การนำเสนอข้อมูล

Graphical Methods for Describing Data

ข้อมูลที่เก็บรวบรวมมาได้ และยังมิได้มีการจัดหมวดหมู่ข้อมูลชนิดนี้เรียกว่า **ข้อมูลดิบ** (Raw Data) ข้อมูลดิบเหล่านี้ ถ้ามีจำนวนมากเราจะไม่สามารถมองเห็นลักษณะของข้อมูลได้จึงต้องมีการจัดเตรียมข้อมูลดิบให้เป็นหมวดหมู่ ถ้าข้อมูลดิบมีจำนวนน้อย ก็ใช้วิธีเรียงข้อมูลจากมากไปหาน้อย หรือจากน้อยไปหามาก ข้อมูลที่เรียงลำดับแบบนี้ เรียกว่า Ungrouped Data แต่ถ้าข้อมูลดิบมีจำนวนมาก ต้องใช้วิธี **การแจกแจงความถี่** (Frequency Distribution)

การแจกแจงความถี่ เป็นการจัดเรียงลำดับข้อมูลดิบที่เก็บรวบรวมมาได้ โดยจัดให้เป็นหมวดหมู่ แล้วหาจำนวนของข้อมูลในแต่ละหมู่ ข้อมูลที่หาได้โดยวิธีการนี้ เรียกว่า ข้อมูลที่เป็นหมวดหมู่ (Grouped Data)

รูปแบบของการแจกแจงความถี่ สามารถทำได้ 2 รูปแบบ คือแบบใช้ตาราง และแบบใช้แผนภูมิหรือกราฟ

การแจกแจงความถี่แบบตาราง

- ตารางแจกแจงความถี่ (Frequency Distribution)
- ตารางแจกแจงความถี่สะสม (Cumulative Frequency Distribution)
- ตารางแจกแจงความถี่สัมพัทธ์ (Relative Frequency Distribution)
- ตารางแจกแจงความถี่สะสมสัมพัทธ์ (Relative Cumulative Frequency Distribution)

การแจกแจงความถี่แบบใช้แผนภูมิหรือกราฟ

- ฮิสโทแกรม (Histogram)
- รูปหลายเหลี่ยมของความถี่ (Frequency Polygon)
- โค้งความถี่ (Frequency Curves)
- โค้งความถี่สะสม (Cumulative Frequency หรือ Ogive Curve)
- แผนภาพต้น-ใบ (Stem-and-Leaf Plot หรือ Stem Plot)

3.1 การแจกแจงความถี่แบบตาราง (Frequency Distribution)

ลักษณะของตารางแจกแจงความถี่โดยทั่วไป ประกอบด้วย ข้อมูล รอยขีด และความถี่ดังตัวอย่างต่อไปนี้

ตัวอย่างที่ 1 ในการทดสอบย่อยครั้งที่ 1 วิชาคณิตศาสตร์ ซึ่งมีคะแนนเต็ม 10 คะแนน มีนักเรียน 10 คน โดยได้คะแนนดังนี้ 5, 5, 0, 9, 2, 7, 8, 10, 8 และ 3 หากนำมาเขียนในรูปตารางแจกแจงความถี่สำหรับทุกค่าของคะแนนที่เป็นไปได้ทั้ง 11 ค่าเป็นดังนี้

x (คะแนน)	0	1	2	3	4	5	6	7	8	9	10
f (ความถี่)											

จากตาราง ตัวแปร คือ...

ค่าจากการสังเกต คือ...

ค่าที่เป็นไปได้ คือ ....

ตัวอย่างที่ 2 คะแนนจากการสอบวิชาคณิตศาสตร์ของนักเรียน 60 คน ซึ่งมีคะแนนเต็ม 30 คะแนน เป็นดังนี้

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

จากข้อมูลข้างต้นสามารถสรุปให้กะทัดรัดได้โดยใช้ตารางแจกแจงความถี่ ดังนี้

คะแนน	**รอยขีด (รอยคะแนน)**	**ความถี่**
16		
17		
18		
19		
20		
21		
22		
23		
24		
25		
26		
27		
28		
29		
รวม		

จากตารางแจกแจงความถี่ข้างต้นทั้ง 2 ตัวอย่าง จะเห็นว่าไม่ค่อยมีประโยชน์ต่อการวิเคราะห์ข้อมูลมากนัก การแจกแจงความถี่ทุกค่าของข้อมูลทำให้เสียเวลา และยากที่จะสรุปผล ดังนั้นแทนที่จะแจกแจงความถี่ของทุกข้อมูล สามารถสร้างตารางแจกแจงความถี่ได้อีกลักษณะหนึ่งโดยการจัดข้อมูลที่ใกล้เคียงกันเป็นหมู่หรือเป็นชั้น (Group or Class) ช่วงกว้างของข้อมูลในแต่ละชั้น เรียกว่า **อันตรภาคชั้น** (Class Interval)

ตัวอย่างที่ 3 คะแนนการทดสอบวิชาคณิตศาสตร์ของนักเรียน 60 คน ซึ่งมีคะแนนเต็ม 100 คะแนน เป็นดังนี้

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

จากตาราง จะได้ว่าข้อมูลชุดนี้มี 9 อันตรภาคชั้น โดยอันตรภาคชั้นที่ 2 มีช่วงคะแนน 50 - 54 เรียกค่าทั้ง 2 นี้ว่าขีดจำกัดล่างและขีดจำกัดบน ในชั้นนี้มีความถี่ หรือจำนวน นักเรียน 4 คน โดยมีคะแนนตั้งแต่ 49.5 ถึง 54.5 คะแนน เราเรียกค่าทั้งสองนี้ว่า **ขอบเขตล่างและขอบเขตบน**

3.1.1 วิธีการสร้างตารางแจกแจงความถี่

1. หาค่าพิสัย โดยคำนวณจาก พิสัย = ข้อมูลสูงสุด - ข้อมูลต่ำสุด
2. กำหนดจำนวนชั้นของข้อมูลตามความเหมาะสม
3. คำนวณหาค่าความกว้างของอันตรภาคชั้น จากสูตร ความกว้างของอันตรภาคชั้นเท่ากับพิสัยหารด้วยจำนวนชั้น กำหนดความกว้างของอันตรภาคชั้นเป็นจำนวนเต็มโดยปัดขึ้นเสมอ (ถ้าหารลงตัวให้บวกเพิ่ม 1 เสมอ)
4. ตีตาราง 3 ช่อง ประกอบด้วย ข้อมูลหรืออันตรภาคชั้น รอยขีดและความถี่
5. เขียนอันตรภาคชั้นของแต่ละชั้น โดยอาศัยความกว้างของอันตรภาคชั้นเป็นเครื่องช่วยจะเริ่มจากชั้นข้อมูลค่าต่ำสุด หรือชั้นข้อมูลค่าสูงสุดก็ได้ โดยมีหลักว่าขีดจำกัดล่างของชั้นข้อมูลค่าต่ำสุดและขีดจำกัดบนของชั้นข้อมูลค่าสูงสุดจะต้องคลุมค่าของข้อมูลทั้งหมด
6. ตรวจดูว่าข้อมูลแต่ละตัวอยู่ในอันตรภาคชั้นใด แล้วขีดรอยขีดให้ตรงกับชั้นนั้นตัวละขีด
7. นับจำนวนรอยขีดใส่ในช่องความถี่

ตัวอย่างที่ 4 คะแนนจากการสอบของนักเรียน 60 คน เป็นดังนี้

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

จงสร้างตารางแจกแจงความถี่ให้มีจำนวนชั้น 5 ชั้น

วิธีทำ 1. พิสัย คือ

2. จำนวนชั้น 5 ชั้น

3. ความกว้างของอันตรภาคชั้น คือ

คะแนน	รอยขีด	จำนวนคน
รวม		

ในการกำหนดจำนวนและความกว้างของอันตรภาคชั้นมีข้อสังเกตดังนี้

1. ความกว้างของแต่ละอันตรภาคชั้นไม่จำเป็นต้องเท่ากันหมด ขึ้นอยู่กับวัตถุประสงค์ของการใช้ข้อมูล
2. ค่าที่สังเกตได้บางค่าอาจต่างไปจากค่าอื่นมาก เช่น ในการสอบครั้งหนึ่งมีผู้สอบได้ 2 คะแนน ในขณะที่คนอื่นได้คะแนน **มากกว่า** 50 คะแนน ควรกำหนดอันตรภาคชั้นแรกเป็นอันตรภาคชั้นเปิด (Open-Ended Class Interval)
3. การกำหนดจำนวนอันตรภาคชั้นในการสร้างตารางแจกแจงความถี่ไม่มีกฎเกณฑ์แน่นอนตายตัว ขึ้นอยู่กับลักษณะการกระจายของข้อมูล รายละเอียดของข้อมูลที่ต้องการทราบด้วย เช่น ถ้าค่าที่สังเกตได้มีความแตกต่างกันมาก มักจะกำหนดให้มีอันตรภาคชั้นน้อย เพื่อไม่ให้มีอันตรภาคชั้นที่มีความถี่เป็นศูนย์ หรือหากต้องการทราบรายละเอียดของข้อมูลอย่างละเอียด ก็ควรกำหนดให้มีจำนวนอันตรภาคชั้นมาก โดยทั่วไปนิยมใช้อยู่ 7-15 ชั้น

ตัวอย่างที่ 5 ตารางแสดงจำนวนผู้ป่วยในเขตตำบลศาลายา พ.ศ. 2554 จำแนกตามระยะเวลาป่วย

ระยะเวลาป่วย (วัน)	จำนวนผู้ป่วย
น้อยกว่า 8 วัน	441,250
8 - 14	50,650
15 - 28	12,560
29 - 42	8,720
มากกว่า 42	22,110
ไม่ทราบข้อมูล	7,850
รวม	**543,140**

จากตัวอย่าง จะเห็นได้ชัดว่าความกว้างของแต่ละอันตรภาคชั้นไม่เท่ากัน และมีอันตรภาคชั้นเปิด

ตัวอย่างที่ 6 จากการสอบถามข้อมูลเกี่ยวกับจำนวนชั่วโมงเฉลี่ยของการทำงานในหนึ่งสัปดาห์ของคนจำนวน 24 คน ดังนี้

35.0	48.0	45.0	43.0	38.2	50.0	39.8	40.7
40.0	50.0	35.4	38.8	40.2	45.0	45.0	40.0
43.0	48.0	43.3	53.1	35.6	41.1	34.8	51.0

จงสร้างตารางแจกแจงความถี่ให้มีจำนวน 5 อันตรภาคชั้น

จำนวนชั่วโมง	รอยขีด	ความถี่
30.0 - 34.9		
รวม		

ตัวอย่างที่ 7 จากตัวอย่างข้างต้น ค่าที่สังเกตได้มีจุดทศนิยมอยู่ด้วย การกำหนดอันตรภาคชั้นอาจกำหนดให้อยู่ในช่วงดังนี้

จำนวนชั่วโมง (x)	ความถี่
$30 \le x < 35$	
รวม	

3.1.2 ตารางแจกแจงความถี่สะสม (Cumulative Frequency Distribution)

ความถี่สะสม (Cumulative Frequency) คือผลรวมของความถี่ของค่านั้นหรือของอันตรภาคชั้นนั้น กับความถี่ของค่า หรือของอันตรภาคชั้นที่มีช่วงคะแนนต่ำกว่าทั้งหมด (หรือสูงกว่าทั้งหมดอย่างใดอย่างหนึ่ง)

ตัวอย่างที่ 8 จากตารางแจกแจงความถี่ของคะแนนการทดสอบวิชาคณิตศาสตร์ของนักเรียน 60 คน ซึ่งมีคะแนนเต็ม 100 คะแนน (ตัวอย่างที่ 3) จงสร้างตารางแจกแจงความถี่สะสม

คะแนน	ความถี่	ความถี่สะสม
45 - 49	2	
50 - 54	4	
55 - 59	8	
60 - 64	10	
65 - 69	15	
70 - 74	11	
75 - 79	6	
80 - 84	3	
85 - 89	1	
รวม	**60**	

3.1.3 ตารางแจกแจงความถี่สัมพัทธ์ (Relative Frequency Distribution)

ความถี่สัมพัทธ์ (Relative Frequency) คืออัตราส่วนระหว่างความถี่ของค่านั้น หรือของอันตรภาคชั้นนั้นกับผลรวมของความถี่ทั้งหมด อาจอยู่ในรูปเศษส่วน ทศนิยมหรือร้อยละ วัตถุประสงค์ในการสร้างเพื่อหาว่าความถี่ของแต่ละค่าที่เป็นไปได้ หรือของแต่ละอันตรภาคชั้นมีจำนวนมากน้อยเพียงใดเมื่อเทียบกับความถี่ทั้งหมด

ตัวอย่างที่ 9 จากตารางแจกแจงความถี่ของคะแนนการทดสอบวิชาคณิตศาสตร์ของนักเรียน 60 คน ซึ่งมีคะแนนเต็ม 100 คะแนน (ตัวอย่างที่ 3) จงสร้างตารางแจกแจงความถี่สัมพัทธ์ และร้อยละของความถี่สัมพัทธ์

คะแนน	ความถี่	ความถี่สัมพัทธ์	ร้อยละของความถี่สัมพัทธ์
45 - 49	2		
50 - 54	4		
55 - 59	8		
60 - 64	10		
65 - 69	15		
70 - 74	11		
75 - 79	6		
80 - 84	3		
85 - 89	1		
รวม	**60**		

3.1.4 ตารางแจกแจงความถี่สะสมสัมพัทธ์ (Relative Cumulative Frequency Distribution)

ความถี่สะสมสัมพัทธ์ (Relative Cumulative Frequency) คืออัตราส่วนระหว่างความถี่สะสมของค่านั้นหรือของอันตรภาคชั้นนั้นกับผลรวมของความถี่ทั้งหมด ซึ่งอาจแสดงในรูปเศษส่วน ทศนิยมหรือร้อยละ ตารางนี้ใช้เพื่อหาว่าแต่ละค่าที่เป็นไปได้หรือแต่ละอันตรภาคชั้นที่มีความถี่สะสมเป็นจำนวนมากน้อยเพียงใดเมื่อเทียบกับความถี่ทั้งหมด

ตัวอย่างที่ 10 จากตารางแจกแจงความถี่ของคะแนนการทดสอบวิชาฟิสิกส์ของนักเรียน 50 คน ซึ่งมีคะแนนเต็ม 50 คะแนน จงสร้างตารางแจกแจงความถี่สะสมสัมพัทธ์ และร้อยละของความถี่สะสมสัมพัทธ์

คะแนน	**ความถี่**	**ความถี่สะสม**	**ความถี่สะสมสัมพัทธ์**	**ร้อยละของความถี่สะสมสัมพัทธ์**
$0 \le x < 10$	5			
$10 \le x < 20$	20			
$20 \le x < 30$	10			
$30 \le x < 40$	8			
$40 \le x < 50$	7			
รวม	**50**			

แบบฝึกทักษะ 3.1

1. ข้อมูลคะแนนสอบวิชาสถิติ ของนักเรียน 36 คน ดังนี้

72	83	82	92	70	91	71	33	42	51	55	75
38	96	85	93	60	75	38	40	75	49	53	41
86	89	51	57	66	92	55	48	85	85	54	56

   a. จงสร้างตารางแจกแจงความถี่ ความถี่สะสม ความถี่สัมพัทธ์ และความถี่สะสมสัมพัทธ์ ที่มีอันตรภาคชั้นเป็น 30-39, 40-49, 50-59, 60-69, 70-79, 80-89, 90-99
   b. จงหาว่าความถี่ในช่วงใดมีค่าสูงสุด
   c. จงหาร้อยละของจำนวนนักเรียนที่ได้คะแนนระหว่าง 70-79 คะแนน
   d. จงหาจำนวนนักเรียนที่ได้คะแนนมากกว่า 59 คะแนน
   e. จงหาร้อยละของนักเรียนที่ได้คะแนนน้อยกว่าหรือเท่ากับ 59 คะแนน

3.2 การแจกแจงความถี่แบบใช้กราฟ

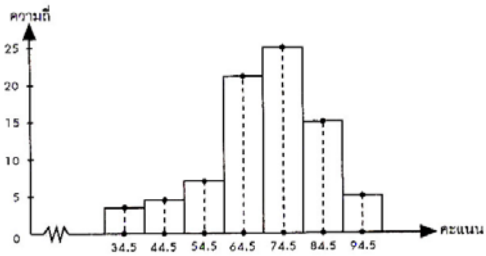
การใช้กราฟแสดงการแจกแจงความถี่ของข้อมูล สามารถทำให้เห็นการกระจายของข้อมูลได้ชัดเจนกว่าการดูจากตารางแจกแจงความถี่ กราฟที่จะกล่าวต่อไปมีดังนี้

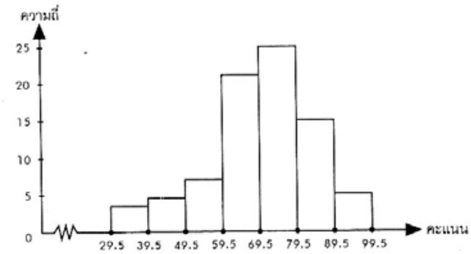
- ฮิสโทแกรม (Histogram)
- รูปหลายเหลี่ยมของความถี่ (Frequency Polygon)
- โค้งความถี่ (Frequency Curves)
- โค้งความถี่สะสม (Cumulative Frequency หรือ Ogive Curve)
- แผนภาพต้น-ใบ (Stem-and-Leaf Plot หรือ Stem Plot)

ฮิสโทแกรม (Histogram) หรือรูปสี่เหลี่ยมผืนผ้าของความถี่ เป็นการแสดงความถี่โดยอาศัยพื้นที่สี่เหลี่ยมผืนผ้าคล้ายกับการนำเสนอข้อมูลแบบกราฟแท่ง โดยที่แต่ละแท่งจะติดกัน เพราะใช้ขอบเขตล่างและขอบเขตบนของข้อมูลแต่ละชั้น มาแบ่งหน่วยบนแกนแนวนอน (แกน X) ส่วนแกนแนวตั้ง (แกน Y) แสดงความถี่ (f) และพื้นที่ของรูปสี่เหลี่ยมผืนผ้าแทนความถี่ของอันตรภาคชั้น

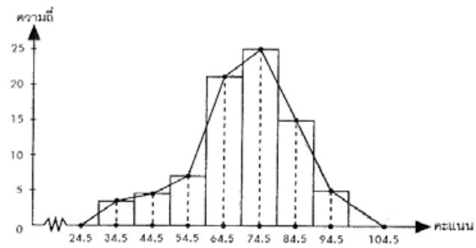
ตัวอย่างที่ 11 จงเขียนฮิสโทแกรมของข้อมูลนี้

คะแนน	ขีดจำกัดชั้น	จำนวน (ความถี่)
30 - 39	29.5 - 39.5	2
40 - 49	39.5 - 49.5	3
50 - 59	49.5 - 59.5	8
60 - 69	59.5 - 69.5	22
70 - 79	69.5 - 79.5	25
80 - 89	79.5 - 89.5	15
90 - 99	89.5 - 99.5	5

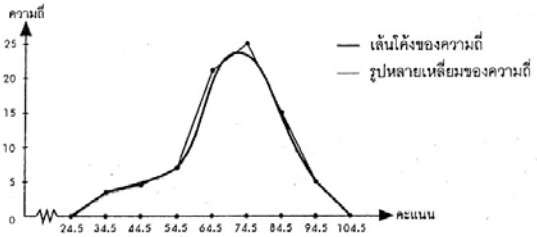




รูปหลายเหลี่ยมของความถี่ (Frequency Polygon) เป็นการแสดงความถี่โดยจุดกึ่งกลางชั้นของข้อมูลและความถี่ของข้อมูลในแต่ละชั้น แบ่งหน่วยบนแกนแนวนอนแสดงจุดกึ่งกลางและแกนแนวตั้งแสดงความถี่ จะได้ตำแหน่งของจุดกึ่งกลางและความถี่ของข้อมูลในแต่ละชั้น เมื่อลากเส้นเชื่อมระหว่างตำแหน่งจากข้อมูลน้อยไปหาข้อมูลมาก ภาพที่ได้เป็นรูปหลายเหลี่ยมของความถี่ หรือสร้างต่อจากการทำฮิสโทแกรมโดยการแบ่งกึ่งกลางที่ยอดของแต่ละแท่งแล้วลากเส้นเชื่อมของจุดแบ่ง จะได้รูปหลายเหลี่ยมของความถี่ตามต้องการ จากตัวอย่างที่ผ่านมาสามารถแสดงข้อมูลโดยใช้รูปหลายเหลี่ยมของความถี่ ได้ดังนี้

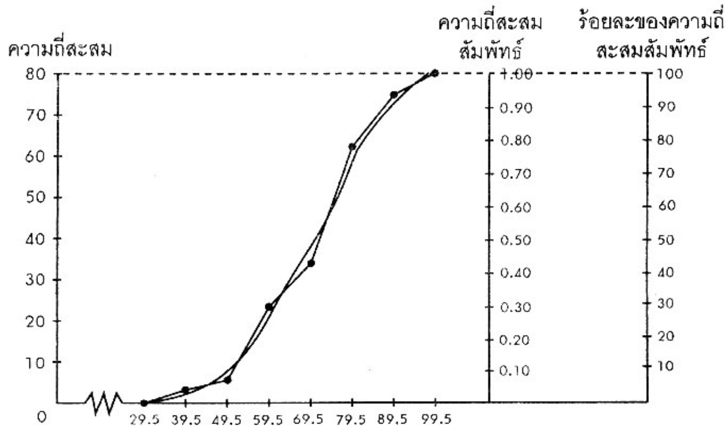


โค้งความถี่ (Frequency Curves) เป็นโค้งที่เกิดจากการปรับเส้นของรูปหลายเหลี่ยมของความถี่ให้เรียบขึ้น โดยการปรับจะต้องให้พื้นที่ภายใต้เส้นโค้งที่ปรับใหม่มีขนาดใกล้เคียงกับพื้นที่ของรูปหลายเหลี่ยมของความถี่



โค้งความถี่สะสม (Cumulative Frequency หรือ Ogive Curve) เป็นโค้งที่แสดงความถี่สะสมของข้อมูลตั้งแต่ค่าต่ำสุด ก่อนสร้างโค้งความถี่สะสม ควรสร้างตารางแจกแจงความถี่สะสมก่อน เมื่อได้ตารางแจกแจงความถี่สะสมแล้วจึงเขียนกราฟความถี่สะสมโดยใช้แกน Y เป็นความถี่สะสมและแกน X เป็นขีดจำกัดชั้นที่แท้จริงของข้อมูลแต่ละชั้น ต่อจากนั้นจึงลากเส้นโยงจุดแต่ละจุดเหล่านั้น

คะแนน	ขีดจำกัดชั้น	จำนวน (ความถี่)	ความถี่สะสม	ความถี่สะสมสัมพัทธ์	ร้อยละของความถี่สะสมสัมพัทธ์
30 - 39	29.5 - 39.5	2	2	0.0250	2.50
40 - 49	39.5 - 49.5	3	5	0.0625	6.20
50 - 59	49.5 - 59.5	8	13	0.1625	16.25
60 - 69	59.5 - 69.5	22	35	0.4375	43.75
70 - 79	69.5 - 79.5	25	60	0.7500	75.00
80 - 89	79.5 - 89.5	15	75	0.9375	93.75
90 - 99	89.5 - 99.5	5	80	1.000	100.00



เทคนิคการเขียนเส้นโค้งของความถี่สะสม

1. ให้แกนนอนเป็นคะแนน และแกนตั้งเป็นความถี่สะสม
2. หาตำแหน่งของจุด (ขอบบน, ความถี่สะสม) ของอันตรภาคชั้นแต่ละชั้น
3. ลากต่อแต่ละจุด แล้วปรับโค้งให้เรียบ

แบบฝึกทักษะ 3.2

1. ข้อมูลต่อไปนี้เป็นความสูงของนักเรียนจำนวน 50 คน

162	165	158	171	169	163	162	165	158	155
154	170	158	158	155	154	152	160	170	159
154	172	162	170	164	170	151	162	160	159
159	157	159	167	160	159	155	172	154	155
173	166	158	156	175	155	165	159	153	163

จงสร้างตารางแจกแจงความถี่โดยให้แต่ละอันตรภาคชั้นมีความกว้างเท่ากัน และมีอันตรภาคชั้น 160-164 ในตารางด้วย สร้างอิสโทแกรม รูปหลายเหลี่ยมแห่งความถี่ และโค้งความถี่ (สร้างในรูปเดียวกันได้โดยเลือกใช้สีให้เห็นชัดเจน)

3.3 แผนภาพต้น-ใบ (Stem-and-Leaf Plot หรือ Stem Plot)

การสร้าง ตารางแจกแจงความถี่และฮีสโทรแกรม ไม่ทำให้สามารถบอกได้ว่าข้อมูลมีค่าใดบ้าง เนื่องจากได้จัดแบ่งข้อมูลที่มีอยู่เป็นช่วงๆ รู้ได้เพียงข้อมูลแต่ละกลุ่มมีมากหรือน้อยเพียงไร การจัดข้อมูลเป็นกลุ่มอาจใช้วิธีการสร้างแผนภาพเพื่อแจกแจงความถี่และวิเคราะห์ข้อมูลเบื้องต้นไปพร้อมๆ กัน ที่เรียกว่า แผนภาพต้น-ใบ

ตัวอย่างที่ 12 ข้อมูลคะแนนสอบวิชาสถิติ ของนักเรียน 36 คน ดังนี้

72	83	82	92	70	91	71	33	42	51	55	75	38	96	85	93	60	75
38	40	75	49	53	41	86	89	51	57	66	92	55	48	85	85	54	56

แบ่งคะแนนเป็น 7 กลุ่ม ดังนี้

กลุ่มที่ 1 คะแนนตั้งแต่ 30 - 39 คะแนน มีดังนี้--

กลุ่มที่ 2 คะแนนตั้งแต่ 40 - 49 คะแนน มีดังนี้--

กลุ่มที่ 3 คะแนนตั้งแต่ 50 - 59 คะแนน มีดังนี้--

กลุ่มที่ 4 คะแนนตั้งแต่ 60 - 69 คะแนน มีดังนี้--

กลุ่มที่ 5 คะแนนตั้งแต่ 70 - 79 คะแนน มีดังนี้--

กลุ่มที่ 6 คะแนนตั้งแต่ 80 - 89 คะแนน มีดังนี้--

กลุ่มที่ 7 คะแนนตั้งแต่ 90 - 99 คะแนน มีดังนี้--

เขียนแผนภาพต้น-ใบได้ดังนี้

ต้น	ใบ
3	--
4	--
5	--
6	--
7	--
8	--
9	--

นอกจากจะใช้แผนภาพต้น-ใบ นำเสนอข้อมูล 1 ชุดดังตัวอย่าง แผนภาพต้นใบยังสามารถนำเสนอข้อมูล 2 ชุดพร้อมกัน และสามารถเปรียบเทียบข้อมูลทั้ง 2 ชุดได้โดยดูจากแผนภาพดังตัวอย่างต่อไปนี้

ตัวอย่างที่ 13 นักเรียนกลุ่ม 10 มีผลการสอบย่อยวิชาคณิตศาสตร์ครั้งที่ 1 และ 2 ซึ่งมีคะแนนเต็ม 100 คะแนนดังนี้

ครั้งที่ 1	40	53	55	58	60	62	65	66	69	70
	72	72	75	75	81	82	85	100	100	100
ครั้งที่ 2	32	39	68	70	75	78	78	78	79	80
	82	84	85	85	85	86	90	93	95	98

เขียนแผนภาพต้น-ใบได้ดังนี้

ใบ (ทดสอบครั้งที่ 1)	ต้น	ใบ (ทดสอบครั้งที่ 2)
	3	------------------------------------
0	4	------------------------------------
8 5 3	5	------------------------------------
9 6 5 2 0	6	------------------------------------
5 5 2 2 0	7	------------------------------------
5 2 1	8	------------------------------------
	9	------------------------------------
0 0 0	10	------------------------------------

แบบฝึกทักษะ 3.3

1. จากข้อมูลความดันโลหิตของครูโรงเรียนมหิดลวิทยานุสรณ์ 30 คน จงสร้างแผนภาพต้นใบ

154	185	170	172	163	165	180	206	155	187
159	183	129	137	135	170	195	174	189	180
151	175	160	135	132	177	168	152	169	144

2. จากการสำรวจเงินที่พกมาทำงานของพนักงานบริษัทแห่งหนึ่งจำนวน 40 คน จงสร้างแผนภาพต้น-ใบ

1000	1520	2000	1360	1540	1480	1360	1880	1950	1470
2000	1840	1250	1240	1750	1690	1350	1420	1950	1640
1860	1980	1530	1350	1380	1480	1470	1260	1940	1360
1860	1950	1420	1120	1320	1180	1190	1660	1970	1350

3. แผนภาพต้น-ใบที่กำหนดให้แสดงค่าที่นักเรียน 30 คนประมาณความสูงของเสาธงของโรงเรียน
   โดยที่ 5 | 4 แทนความสูง 5.4 เมตร

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

a. มีนักเรียนทั้งหมดกี่คนที่ทำการประมาณ ---------------------------------

b. ค่าประมาณของเสาธงที่สูงที่สุด มีความสูงเท่าไร ---------------------------------

c. ถ้าเสาธงสูงจริง 4 เมตร มีนักเรียนร้อยละเท่าใดที่ประมาณผิด ---------------------------------

4. การวัดค่ากลางของข้อมูล

Measures of Central Value

▶ **ยังจำได้ไหม**
- **ประชากร** (Population) หมายถึงกลุ่มของสมาชิกทุกหน่วยที่เราต้องการศึกษาลักษณะ
- **พารามิเตอร์** (Parameter) หมายถึง ตัวเลขซึ่งแสดงคุณสมบัติบางประการของประชากร เช่น μ, σ^2, σ เป็นต้น
- **ตัวอย่าง** (Sample) หมายถึง กลุ่มย่อยของสมาชิกในกลุ่มประชากรที่เลือกมาเพื่อศึกษาลักษณะที่สนใจ
- **ค่าสถิติ** (Statistic) หมายถึง ตัวเลขที่วัดผลที่ได้จากตัวอย่าง เช่น $\bar{x}$, s^2 , s เป็นต้น

การหาค่ากลางของข้อมูลเพื่อหาค่าสถิติหรือค่าพารามิเตอร์ แล้วนำผลที่ได้มาสรุปและตีความหมายของข้อมูล ใช้เป็นตัวแทนของข้อมูลทั้งหมดเพื่อความสะดวกในการสรุปเรื่องราวเกี่ยวกับข้อมูลนั้นๆ จะช่วยทำให้เกิดการวิเคราะห์ข้อมูลถูกต้องดีขึ้น การหาค่ากลางของข้อมูลมีวิธีหาหลายวิธี แต่ละวิธีมีข้อดีและข้อเสีย และมีความเหมาะสมในการนำไปใช้ไม่เหมือนกัน ขึ้นอยู่กับลักษณะข้อมูลและวัตถุประสงค์ของผู้ใช้ข้อมูลนั้นๆ

ค่าวัดแนวโน้มสู่ส่วนกลางหรือค่ากลางที่เป็นตัวแทนของข้อมูลที่นิยมใช้มีอยู่ 3 ค่า ได้แก่ ค่าเฉลี่ยเลขคณิต มัธยฐานและฐานนิยม

4.1 ค่าเฉลี่ยเลขคณิต (arithmetic mean)

ค่าเฉลี่ยเลขคณิต (arithmetic mean) คือค่าของผลรวมของค่าสังเกตของข้อมูลทั้งหมด หารด้วยจำนวนของข้อมูลทั้งหมด เรียกสั้นๆ ว่าค่าเฉลี่ย ค่าเฉลี่ยเลขคณิตเหมาะที่จะนำมาเป็นค่ากลางของข้อมูลเมื่อข้อมูลนั้นไม่มีค่าใดค่าหนึ่งสูงหรือต่ำผิดปกติ มีสูตรดังนี้

(สำหรับข้อมูลที่ไม่แจกแจงความถี่)

ค่าเฉลี่ยประชากร (population mean)	**ค่าเฉลี่ยกลุ่มตัวอย่าง** (sample mean)
$\mu = \dfrac{\sum_{i=1}^{N} x_i}{N}$	$\bar{x} = \dfrac{\sum_{i=1}^{n} x_i}{n}$

ตัวอย่างที่ 14 คะแนนผลสัมฤทธิ์ทางการเรียนวิชาสถิติเพื่อการวิจัยของนักเรียนกลุ่มตัวอย่างจำนวน 10 คน มีค่าดังนี้

87 61 75 85 73 65 58 66 78 95

จงหาค่าเฉลี่ยของคะแนนผลสัมฤทธิ์ทางการเรียนของนักเรียนกลุ่มนี้

วิธีทำ จากสูตร $\bar{x} = \dfrac{\sum_{i=1}^{n} x_i}{n}$ (**Note:** ให้นักเรียนเขียนสูตรก่อนเสมอ)

จะได้

ตัวอย่างที่ 15 ในการสอบวิชาสถิติของนักเรียนห้องหนึ่งค่าเฉลี่ยเลขคณิตของคะแนนเท่ากับ 53 คะแนน แต่จากการตรวจสอบพบว่า มีข้อสอบของนักเรียน 2 คนที่ยังไม่ได้ตรวจ เมื่อตรวจเสร็จคำนวณค่าเฉลี่ยใหม่ได้ 55 คะแนน และผลรวมของคะแนนสอบเพิ่มขึ้นอีก 180 คะแนน จำนวนนักเรียนในห้องนี้มีเท่าใด

(ข้อนี้เราสนใจคะแนนสถิติของนักเรียนห้องนี้ นั่นคือประชากรคือนักเรียนในห้องนี้)

ตัวอย่างที่ 16 นักเรียนกลุ่มตัวอย่างมี 10 คน มีคะแนนเฉลี่ยวิชาคณิตศาสตร์ 45 คะแนน ต่อมาทราบว่าคิดคะแนนผิดไป 2 คน คือจาก 48 และ 50 คะแนน คิดเป็น 43 และ 60 คะแนน จงหาคะแนนเฉลี่ยที่ถูกต้อง

ตัวอย่างที่ 17 ถ้าผู้สอนจะให้เกรด 4 แก่นักเรียนที่ได้คะแนนไม่ต่ำกว่า 75 คะแนน จากการสอบทั้ง 6 ครั้ง ถ้าคะแนนเฉลี่ยของการสอบย่อย 5 ครั้งของบอลเท่ากับ 71 คะแนน จงหาว่าครั้งที่ 6 บอลต้องสอบได้กี่คะแนนจึงจะได้เกรด 4

ค่าเฉลี่ยเลขคณิตถ่วงน้ำหนัก (weight arithmetic mean)

ค่าเฉลี่ยเลขคณิตถ่วงน้ำหนัก (weight arithmetic mean) ใช้ในกรณีที่ข้อมูลมีความสำคัญไม่เท่ากัน เช่นการหาผลการเรียนเฉลี่ย เนื่องจากแต่ละวิชามีจำนวนหน่วยกิตไม่เท่ากันจึงจำเป็นต้องถ่วงน้ำหนัก

ถ้าให้ $w_1, w_2, \ldots, w_N$ เป็นน้ำหนักถ่วงของค่าสังเกต ค่าเฉลี่ยเลขคณิตถ่วงน้ำหนัก มีสูตรดังนี้

(สำหรับข้อมูลที่ไม่แจกแจงความถี่) **ค่าเฉลี่ยถ่วงน้ำหนัก** $\mu = \dfrac{\sum_{i=1}^{N} w_i x_i}{\sum_{i=1}^{N} w_i}$

ตัวอย่างที่ 18 จงคำนวณหาผลการเรียนเฉลี่ยของ นักเรียนคนหนึ่งซึ่งมีผลการเรียนดังนี้

วิชาที่	**คณิตศาสตร์**	**ชีววิทยา**	**เคมี**	**ฟิสิกส์**	**สังคม**
หน่วยกิต	3	3	2	2	1
เกรด	A	B	B	A	A

ตัวอย่างที่ 19 กำหนดให้ X เป็นตัวแปรหนึ่ง ถ้าค่าที่สังเกตได้พร้อมกับร้อยละของความถี่สะสมมีค่าดังตาราง จงหาค่าเฉลี่ยเลขคณิตของข้อมูล

ค่าที่สังเกต	- 4	- 3	1	2	3
ความถี่สะสม	30	50	60	80	100

ค่าเฉลี่ยเลขคณิตรวม (combined arithmetic mean)

ในการวิเคราะห์ข้อมูลหลาย ๆ ชุดที่หาค่าเฉลี่ยไว้แล้ว หากต้องการหาค่าเฉลี่ยเลขคณิตของข้อมูลทั้งหมดโดยนับรวมเป็นชุดเดียว ต้องใช้การคำนวณโดยค่าเฉลี่ยเลขคณิตรวม

ถ้า $\overline{x}_1, \overline{x}_2, \overline{x}_3, \ldots, \overline{x}_k$ เป็นค่าเฉลี่ยเลขคณิตของข้อมูลชุดที่ 1, 2, 3, ..., k และ

$n_1, n_2, n_3, \ldots, n_k$ เป็นจำนวนค่าจากการสังเกตในข้อมูลชุดที่ 1, 2, 3, ..., k

ค่าเฉลี่ยเลขคณิตรวม $$\overline{x} = \frac{\sum_{i=1}^{k} n_i \overline{x}_i}{\sum_{i=1}^{k} n_i}$$ (ข้อมูลระดับประชากรยังคงใช้สูตรทำนองเดียวกัน)

ตัวอย่างที่ 20 นักเรียนห้องหนึ่งมีนักเรียนชาย 13 คน หญิง 11 คน นักเรียนชายมีความสูงเฉลี่ย 168 เซนติเมตร นักเรียนหญิงมีความสูงเฉลี่ย 155 เซนติเมตร จงหาค่าเฉลี่ยความสูงของนักเรียนทั้งห้อง

ค่าเฉลี่ยเลขคณิตของข้อมูลที่แจกแจงความถี่แล้ว

การคำนวณหาค่าเฉลี่ยเลขคณิตโดยวิธีนี้ใช้สูตรทำนองเดียวกับการหาค่าเฉลี่ยโดยวิธีถ่วงน้ำหนัก โดยที่ความสำคัญของน้ำหนักในที่นี้คือความถี่ของค่าจากการสังเกตแต่ละค่า หรือค่าที่เป็นตัวแทนของแต่ละอันตรภาคชั้น ซึ่งเรียกว่าจุดกึ่งกลางของอันตรภาคชั้น (midpoint)

ค่าเฉลี่ยเลขคณิตของข้อมูลที่แจกแจงความถี่แล้ว $$\overline{x} = \frac{\sum_{i=1}^{k} f_i x_i}{n}$$ (ข้อมูลระดับประชากรยังคงใช้สูตรทำนองเดียวกัน)

เมื่อ k คือจำนวนอันตรภาคชั้น และ x_i เป็นจุดกึ่งกลางชั้นที่ i

ตัวอย่างที่ 21 จากตารางแจกแจงความถี่ จงหาค่าเฉลี่ยเลขคณิตของข้อมูลกลุ่มตัวอย่างต่อไปนี้

ช่วงคะแนน	**จุดกึ่งกลาง (x_i)**	**ความถี่**	$f_i x_i$
0 - 4		3	
5 - 9		4	
10 - 14		10	
15 - 19		2	
20 - 24		1	
รวม			

ค่าเฉลี่ยเลขคณิตของข้อมูล คือ

ตัวอย่างที่ 22 จากตารางแจกแจงความถี่แสดงเงินเดือนของพนักงาน จงหาค่าเฉลี่ยเลขคณิตของข้อมูลประชากรกลุ่มนี้

เงินเดือน	**จำนวน**
6500 - 6999	10
7000 - 7499	15
7500 - 7999	20
8000 - 8499	15
8500 - 8999	10
9000 - 9499	3
9500 - 9999	2

เทคนิคคิดลัด

ค่าเฉลี่ยเลขคณิตโดยวิธีนี้ใช้สูตรลดทอนดังนี้

$$\overline{x} = a + I\,\frac{\sum_{i=1}^{k} f_i d_i}{n} \text{ เมื่อ } d_i = \frac{x_i - a}{I} \text{ และ } k \text{ เป็นจำนวนอันตรภาคชั้น}$$

โดยกำหนดให้ a เป็นค่ากลางสมมุติ โดยค่านี้ได้จากการเลือกจากจุดกึ่งกลางของชั้นใดก็ได้ แต่นิยมใช้ชั้นที่มีความถี่สูงสุด หรือชั้นที่อยู่ตรงกลาง

เมื่อ I แทนความกว้างของอันตรภาคชั้น

d_i แทนจุดกึ่งกลางใหม่ของแต่ละอันตรภาคชั้น

f_i แทนความถี่ของแต่ละอันตรภาคชั้น

n แทนจำนวนข้อมูลทั้งหมด

ตัวอย่างที่ 23 จากตารางแจกแจงความถี่อายุการใช้งานของหลอดไฟ 40 ดวง จงหาค่าเฉลี่ยเลขคณิตของอายุหลอดไฟ

อายุ (ชั่วโมง)	**จำนวน** (f_i)	x_i	$d_i = \frac{x_i - a}{I}$	$f_i\ d_i$
118 - 122	2			
123 - 127	8			
128 - 132	15			
133 - 137	11			
138 - 142	3			
143 - 147	1			

4.2 ค่าเฉลี่ยฮาร์โมนิก (Harmonic Mean หรือ H.M.)

กำหนดให้ $x_1, x_2, x_3, \dots, x_n$ เป็นข้อมูล n จำนวน ซึ่งเป็นค่าบวก ค่าเฉลี่ยฮาร์โมนิกมีสูตรดังนี้

(สำหรับกรณีที่ข้อมูลที่ไม่แจกแจงความถี่)	(สำหรับกรณีที่ข้อมูลแจกแจงความถี่)
$\overline{x} = \dfrac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3} + \dots + \frac{1}{x_n}} = \dfrac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$	$\overline{x} = \dfrac{f_1 + f_2 + f_3 + \dots + f_k}{\frac{f_1}{x_1} + \frac{f_2}{x_2} + \frac{f_3}{x_3} + \dots + \frac{f_k}{x_k}} = \dfrac{\sum_{i=1}^{k} f_i}{\sum_{i=1}^{k} \frac{f_i}{x_i}}$

ข้อมูลระดับประชากรยังคงใช้สูตรทำนองเดียวกัน

ตัวอย่างที่ 24 กำหนดข้อมูล 5, 3, 2 จงหาค่าเฉลี่ยฮาร์โมนิก

ตัวอย่างที่ 25 บอลวิ่งรอบสนามรูปสี่เหลี่ยมจัตุรัส โดยด้านแรกวิ่งด้วยความเร็ว 10 เมตรต่อวินาที เขาเริ่มเหนื่อยจึงวิ่งช้าลงเป็น 8, 7 และ 5 เมตรต่อวินาที ในด้านที่ 2, 3 และ 4 ตามลำดับ จงหาอัตาเร็วเฉลี่ยในการวิ่งของบอล

ตัวอย่างที่ 26 บ้านเมี้ยนกับเตยอยู่ห่างกัน 50 กม. ถ้าเมี้ยนเดินทางไปหาเตยโดยที่ 25 กม. แรกเดินทางด้วยอัตราเร็ว 9 กม./ชม. และ 25 กม. หลังเดินทางด้วยอัตราเร็ว 7 กม./ชม. จงหาอัตราเร็วเฉลี่ยในการเดินทาง

4.3 ค่าเฉลี่ยเรขาคณิต (Geometric Mean)

ค่าค่าเฉลี่ยเรขาคณิตมีประโยชน์เมื่อมีค่าของข้อมูลสูงหรือต่ำกว่าค่าอื่นๆ รวมอยู่ ให้ x_i เป็นข้อมูลซึ่งเป็นค่าบวกและไม่มีจำนวนใดมีค่า 0 ค่าเฉลี่ยเรขาคณิต มีสูตรดังนี้

(สำหรับกรณีที่ข้อมูลที่ไม่แจกแจงความถี่)	(สำหรับกรณีที่ข้อมูลแจกแจงความถี่)
$G = \sqrt[n]{x_1 \cdot x_2 \cdot x_3 \cdot \ldots \cdot x_n}$	$G = \sqrt[n]{x_1^{f_1} \cdot x_2^{f_2} \cdot x_3^{f_3} \cdot \ldots \cdot x_k^{f_k}}$

ข้อมูลระดับประชากรยังคงใช้สูตรทำนองเดียวกัน

เราสามารถใช้ลอการิทึมช่วยในการหาค่าเฉลี่ยเรขาคณิต ดังนี้

(สำหรับกรณีที่ข้อมูลที่ไม่แจกแจงความถี่)	(สำหรับกรณีที่ข้อมูลแจกแจงความถี่)
$\log G = \frac{1}{n}\sum_{i=1}^{n} \log x_i$	$\log G = \frac{1}{n}\sum_{i=1}^{k} f_i \log x_i$

ข้อมูลระดับประชากรยังคงใช้สูตรทำนองเดียวกัน **Note:** $\sum_{i=1}^{k} f_i = n$

ตัวอย่างที่ 27 จงหาค่าเฉลี่ยเรขาคณิตของข้อมูล 2, 4, 8, 16, 32

ตัวอย่างที่ 28 จงหาค่าเฉลี่ยเรขาคณิตของข้อมูล 2, 4, 4, 8

4.4 มัธยฐาน

มัธยฐาน (Median : Me) คือ ค่าที่อยู่แหน่งกึ่งกลางของข้อมูลทั้งชุดเมื่อมีการจัดเรียงคะแนนตามความมากน้อย แบ่งข้อมูลที่เรียงลำดับแล้วออกเป็น 2 ส่วนเท่าๆ กัน ใช้กับข้อมูลเชิงปริมาณเท่านั้น เหมาะที่จะนำมาใช้เป็นค่ากลาง เมื่อข้อมูลนั้นมีค่าหนึ่งค่าใดสูงหรือต่ำผิดปกติ

ค่ามัธยฐานของข้อมูลที่ไม่ได้แจกแจงความถี่ คือค่าของข้อมูลตำแหน่งที่ $\frac{N+1}{2}$

ขั้นตอนการหาค่ามัธยฐาน มีดังนี้

1. จัดเรียงคะแนนความมากน้อย
2. หาตำแหน่งกึ่งกลางของชุดข้อมูล โดยใช้สูตร
   ตำแหน่งของคะแนนกึ่งกลาง $\frac{N+1}{2}$ เมื่อ N แทนจำนวนคะแนนในชุดข้อมูล
3. หาค่ามัธยฐาน โดยการอ่านค่าคะแนน ณ ตำแหน่งที่คำนวณได้ในตอนที่ 2
   นั้นคือ Me = $x_{(\frac{N+1}{2})}$

ตัวอย่างที่ 29 จงหาค่ามัธยฐาน

ก. 2 5 1 4 6 7 9

ข. 2 5 1 4 6 7 9 10

ค. 2 5 1 4 6 7 9 10 8

ตัวอย่างที่ 30 จงหาค่ามัธยฐานของจำนวนเงินฝากในรอบ 8 ปีของธนาคารแห่งหนึ่ง

พ.ศ.	จำนวน (ล้านล้านบาท)
2547	2.43
2548	2.76
2549	3.25
2550	3.68
2551	4.31
2552	4.96
2553	4.67
2554	3.97

การหามัธยฐานของข้อมูลที่แจกแจงความถี่

หามัธยฐานของข้อมูลที่แจกแจงความถี่ สามารถหาได้ดังนี้

1. โดยใช้การเปรียบเทียบสัดส่วน
2. โดยใช้สูตร ดังนี้

   ค่ามัธยฐานของข้อมูลที่แจกแจงความถี่ คือ $L + I\left(\frac{\frac{N}{2} - F}{f_M}\right)$

   เมื่อ L คือค่าขอบเขตล่างของชั้นที่มีมัธยฐานอยู่

   I คือความกว้างของอันตรภาคชั้นที่มีมัธยฐานอยู่

   F คือความถี่สะสมของชั้นที่อยู่ก่อนชั้นที่มีมัธยฐาน

   f_m ความถี่ของชั้นที่มีมัธยฐาน
3. โดยใช้กราฟความถี่สะสม (เป็นการประมาณที่หยาบกว่า)

ตัวอย่างที่ 31 จงหาค่ามัธยฐาน

คะแนน	จำนวนนักเรียน	ช่วงคะแนนที่แท้จริง	ความถี่สะสม
30 – 39	8		
40 – 49	10		
50 – 59	12		
60 – 69	45		
70 – 79	50		
80 – 89	20		
90 – 99	15		

สมบัติของมัธยฐาน

ผลรวมของค่าสัมบูรณ์ของผลต่างระหว่างข้อมูลแต่ละค่ากับค่ามัธยฐานของข้อมูลชุดนั้นจะมีค่าน้อยที่สุด

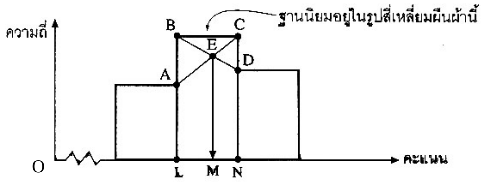
กล่าวคือ $\sum_{i=1}^{N} |x_i - Me|$ มีค่าน้อยที่สุด

4.5 ฐานนิยม (Mode)

ฐานนิยม คือค่าสังเกตที่เกิดซ้ำมากที่สุดหรือมีความถี่สูงสุด จะใช้กับข้อมูลเชิงคุณภาพมากกว่าเชิงปริมาณ เช่น ขนาดรองเท้า อายุ ความสูง ถ้าข้อมูลไม่ซ้ำกันเลยถือว่าไม่มีฐานนิยม ข้อมูลชุดหนึ่งอาจมีฐานนิยมมากกว่าหนึ่งค่าก็ได้ กรณีที่ข้อมูลใดมีฐานนิยมมากกว่า 2 ค่า อาจถือได้ว่าข้อมูลชุดนั้นไม่มีฐานนิยมเลยก็ได้

สำหรับข้อมูลที่ไม่แจกแจงความถี่ ฐานนิยมคือข้อมูลตัวที่ซ้ำกันมากที่สุด
สำหรับข้อมูลที่แจกแจงความถี่

- ให้จุดกึ่งกลางชั้นที่มีความถี่สูงสุดเป็นค่าประมาณของฐานนิยม หรือ
- หาจากฮิสโทแกรม กำหนดให้ $Mode$ คือค่าฐานนิยม



จากรูปจะได้ $Mode = LO + ML$

ให้ $AB = d_1, CD = d_2$ และ I คือความกว้างของชั้นที่มีฐานนิยมอยู่

จะได้ $MN = I - ML$

จากสามเหลี่ยมคล้ายจะได้ $\frac{AB}{CD} = \frac{ML}{MN}$

แทนค่า $\frac{d_1}{d_2} = \frac{ML}{\mathrm{I}-ML}$

จะได้ $ML = \left(\frac{d_1}{d_1+d_2}\right)I$

จาก $Mode = LO + ML$

จะได้ $Mode = LO + I\left(\frac{d_1}{d_1+d_2}\right)$

ให้ LO เป็นค่าขอบเขตล่างของชั้นที่มีฐานนิยมอยู่ แทนด้วย L

d_1, d_2 เป็นผลต่างของความถี่ของชั้นที่มีฐานนิยมอยู่กับความถี่ของขั้นที่ติดกัน ซึ่งเป็นช่วงคะแนนที่ต่ำกว่าและสูงกว่าตามลำดับ

จะได้สูตรดังนี้

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

ตัวอย่างที่ 32 ผลการสอบของนักเรียน 10 คนเป็นดังนี้ 15 20 15 9 18 14 12 15 7 6
จงหาค่าเฉลี่ยเลขคณิต มัธยฐาน และฐานนิยม

ตัวอย่างที่ 33 จากตารางแจกแจงความถี่ จงสร้างฮิสโทรแกรม และหาฐานนิยม

คะแนน	จำนวน
10 − 19	3
20 − 29	8
30 − 39	8
40 − 49	5
50 − 59	2

ตัวอย่างที่ 34 จากตารางแจกแจงความถี่ จงหาฐานนิยม

คะแนน	จำนวน
10 − 19	3
20 − 29	8
30 − 39	12

กรณีความกว้างของแต่ละอันตรภาคชั้นไม่เท่ากัน

จะต้องดูจากอัตราส่วนระหว่างความถี่ต่อความกว้างของอันตรภาคชั้น ดังตัวอย่างต่อไปนี้

ตัวอย่างที่ 35 จากตารางแจกแจงความถี่ จงหาฐานนิยม

คะแนน	จำนวน	ช่วงคะแนนที่แท้จริง	ความกว้างของอันตรภาคชั้น	$\frac{f}{I}$
5 − 7	6			
8 − 14	28			
15 − 24	30			
25 − 28	4			

แบบฝึกทักษะ 4

1. ตารางต่อไปนี้เป็นตัวเลขเงินเดือนของพนักงาน 100 คน ในบริษัทแห่งหนึ่ง

เงินเดือนไม่ต่ำกว่า (บาท)	จำนวน
3,000	100
4,000	65
5,000	30
6,000	14
7,000	7
8,000	4
9,000	2

จงเติมตารางต่อไปนี้ให้สมบูรณ์เพื่อแสดงการแจกแจงความถี่ และหาค่าเฉลี่ยเลขคณิต มัธยฐาน และฐานนิยม (ถ้ามี)

เงินเดือน	จำนวน	ขอบล่าง	ขอบบน	จุดกึ่งกลางชั้น	ความถี่สะสม	ร้อยละของความถี่สัมพัทธ์	ร้อยละของความถี่สะสมสัมพัทธ์
3,000 - 3,999	35						
4,000 - 4,999							
9,000 ขึ้นไป							
รวม							

2. ผลการสอบของดาว 4 วิชาเป็นดังนี้ 85, 89, 87 และ 96 คะแนน ถ้าการสอบครั้งนี้มี 5 วิชา และดาวคาดหวังว่าจะได้ค่าเฉลี่ย 90 คะแนนเป็นอย่างน้อย จงหาว่าวิชาที่ 5 เธอต้องได้คะแนนน้อยสุดเท่าไรจึงจะเป็นดังหวัง

3. ข้อมูลชุดหนึ่งประกอบด้วย 17, 14, 11, 6 และ x จงหาค่าของ x ที่ทำให้ค่าเฉลี่ยเลขคณิตและมัธยฐานของขั้มูลชุดนี้มีค่าเท่ากัน

4. ถ้าอุณหภูมิของแต่ละวันใน 1 สัปดาห์เป็นดังนี้ 32, 36, 35, 34, 37, 31 และ 34 องศาเซลเซียส จงหาค่าเฉลี่ยเลขคณิตของอุณหภูมิที่มีหน่วยเป็นองศาฟาเรนไฮต์

5. แผนภาพต้น-ใบของข้อมูลชุดหนึ่งเป็นดังนี้

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

จากข้อมูลข้างต้น
1. จงหาค่าเฉลี่ยเลขคณิต มัธยฐาน และฐานนิยมของข้อมูลชุดนี้
2. จงพิจารณาว่าควรใช้ค่ากลางชนิดใดเพื่อเป็นตัวแทนข้อมูล พร้อมทั้งให้เหตุผลประกอบ
3. ข้อมูลที่มากกว่า 40 ร้อยละเท่าไรของข้อมูลทั้งหมด

5. การวัดตำแหน่งที่ของข้อมูล

Measures of Relative Standing

การวัดตำแหน่ง เป็นการแปลงข้อมูลแต่ละชุดให้อยู่ในลักษณะเดียวกัน เพื่อประโยชน์ในการเปรียบเทียบข้อมูลระหว่างข้อมูลคนละชุดกัน การแปลงข้อมูลมีลักษณะเป็นการแบ่งชุดข้อมูลออกเป็นส่วนย่อยๆ มีทั้งแบ่งออกเป็น 4, 10 และ 100 ส่วน

การวัดตำแหน่งที่แบ่งข้อมูลออกเป็น 4 ส่วน คือ **ควอไทล์**
การวัดตำแหน่งที่แบ่งข้อมูลออกเป็น 10 ส่วน คือ **เดไซล์**
การวัดตำแหน่งที่แบ่งข้อมูลออกเป็น 100 ส่วน คือ **เปอร์เซ็นไทล์**

ข้อมูลแต่ละชุดมีลักษณะแตกต่างกัน ดังนั้นการจะนำคะแนนที่อยู่ต่างชุดกันมาเปรียบเทียบกันจึงจำเป็นต้องนำข้อมูลแต่ละชุดนั้นมาแปลงให้มีลักษณะเดียวกันเสียก่อน โดยอาจใช้วิธีการทางสถิติชนิดใดชนิดหนึ่งคือ ควอไทล์ เดไซล์ หรือเปอร์เซ็นไทล์

5.1 การหา Q_k, D_k และ P_k สำหรับข้อมูลที่ไม่แจกแจงความถี่

ขั้นตอนการหาค่าของคะแนน ณ ตำแหน่งควอไทล์ เดไซล์ และเปอร์เซ็นไทล์ ที่กำหนด

1. จัดเรียงคะแนนทั้งหมดในชุดข้อมูล โดยเรียงจากค่าน้อยไปหาค่ามาก
2. คำนวณหาควอไทล์ เดไซล์ และเปอร์เซ็นไทล์ จากสูตร

$$\boxed{\begin{aligned} Q_k &= x_{\frac{k(N+1)}{4}} \\ D_k &= x_{\frac{k(N+1)}{10}} \\ P_k &= x_{\frac{k(N+1)}{100}} \end{aligned}}$$

(ข้อมูลระดับตัวอย่างยังคงใช้สูตรทำนองเดียวกัน)

โดยที่ k คือตำแหน่งของควอไทล์ เดไซล์ หรือเปอร์เซ็นไทล์
x คือข้อมูล
n คือจำนวนของข้อมูลทั้งหมด

และ x_1 คือข้อมูลตำแหน่งที่ 1
x_2 คือข้อมูลตำแหน่งที่ 2
$\vdots$
x_N คือข้อมูลตำแหน่งที่ N

ตัวอย่างที่ 36 จากการทดสอบระดับสติปัญญาของเรียน 10 คน ปรากฏผลดังนี้
105 120 86 125 135 82 102 115 132 92
จงหาควอไทล์ที่ 1, 2 และ 3

ตัวอย่างที่ 37 จากข้อมูลในตัวอย่างที่ 36 จงหา D_7 , P_{30} และ P_{75}

5.2 การหา Q_k, D_k และ P_k สำหรับข้อมูลที่แจกแจงความถี่

ขั้นตอนการหาค่าของคะแนน ณ ตำแหน่งควอไทล์ เดไซล์ และเปอร์เซ็นไทล์ ที่กำหนด

1. สร้างความถี่สะสม
2. คำนวณหาควอไทล์ เดไซล์ และเปอร์เซ็นไทล์ จากสูตร

$$Q_k = x_{\frac{kN}{4}}$$

$$D_k = x_{\frac{kN}{10}}$$

$$P_k = x_{\frac{kN}{100}}$$

(ข้อมูลระดับตัวอย่างยังคงใช้สูตรทำนองเดียวกัน)

โดยที่ k คือตำแหน่งของควอไทล์ เดไซล์ หรือเปอร์เซ็นไทล์

x คือข้อมูล

N คือจำนวนของข้อมูลทั้งหมด

และ x_1 คือข้อมูลตำแหน่งที่ 1

x_2 คือข้อมูลตำแหน่งที่ 2

⋮

x_N คือข้อมูลตำแหน่งที่ N

3. หรือคำนวณหาควอไทล์ เดไซล์ และเปอร์เซ็นไทล์ โดยการเทียบอัตราส่วนของความแตกต่างของความถี่สะสมกับความแตกต่างของค่า ใช้สูตรดังนี้

$$Q_k = L + I\left(\frac{\frac{kN}{4} - F}{f_M}\right)$$

$$D_k = L + I\left(\frac{\frac{kN}{10} - F}{f_M}\right)$$

$$P_k = L + I\left(\frac{\frac{kN}{100} - F}{f_M}\right)$$

(ข้อมูลระดับตัวอย่างยังคงใช้สูตรทำนองเดียวกัน)

เมื่อ L คือค่าขอบเขตล่างของชั้นที่มีควอไทล์ เดไซล์ หรือเปอร์เซ็นไทล์

I คือความกว้างของอันตรภาคชั้น

$\frac{kN}{4}, \frac{kN}{10}, \frac{kN}{100}$ คือตำแหน่งของควอไทล์ เดไซล์ หรือเปอร์เซ็นไทล์

F คือความถี่สะสมของชั้นที่อยู่ก่อนชั้นที่มีควอไทล์ เดไซล์ หรือเปอร์เซ็นไทล์

f_M ความถี่ของชั้นที่มีควอไทล์ เดไซล์ หรือเปอร์เซ็นไทล์

ตัวอย่างที่ 38 ผลการสอบวิชาคณิตศาสตร์ของนักเรียน 47 คน ปรากฏผลดังตาราง

คะแนน	**ความถี่**
15 – 24	3
25 – 34	1
35 – 44	6
45 – 54	8
55 – 64	19
65 – 74	2
75 – 84	7
85 – 94	1

จงหา

1. นักเรียนที่มีคะแนนต่ำกว่าประมาณหนึ่งในสี่ของนักเรียนทั้ง 47 คน มีคะแนนเท่าใด
2. ถ้านำคะแนนที่ได้มาเปลี่ยนเป็นระดับเกรด A, B, C, D และ E โดยให้มีผู้ได้ระดับคะแนน A เป็นจำนวน 10% ผู้ที่ได้ระดับคะแนน A จะต้องมีคะแนนเท่าใด

ตัวอย่างที่ 39 กำหนดข้อมูล 15 50 4 20 7 30 35 48 24

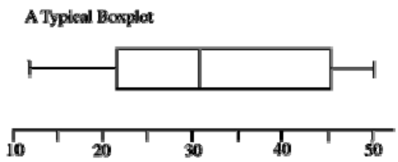
จงหาว่า 1. ค่า 20 ตรงกับเปอร์เซ็นไทล์ที่เท่าไร

2. ค่า 10 ตรงกับเปอร์เซ็นไทล์ที่เท่าไร

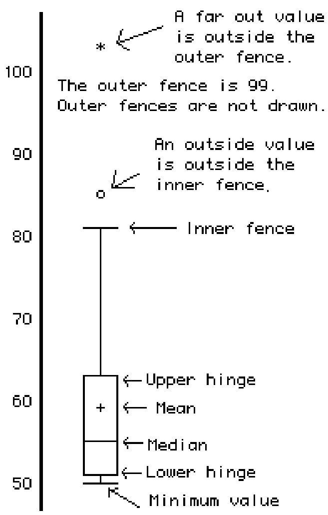
5.3 บอกซ์พลอต (Box Plots)

บอกซ์พลอตเป็นการสรุปลักษณะของข้อมูลชุดหนึ่ง ๆ ด้วยกราฟ ในการสร้างบอกซ์พลอตเราต้องทราบค่าสถิติทั้งหมด 5 ค่า คือค่าต่ำสุดของข้อมูล ค่าควอไทล์ที่ 1 (เปอร์เซนต์ไทล์ที่ 25) ค่ามัธยฐาน ค่าควอไทล์ที่ 3 (เปอร์เซนต์ไทล์ที่ 75) และค่าสูงสุดของข้อมูล ขั้นตอนในการสร้างบอกซ์พลอตมีดังนี้

1. สร้างกล่องโดยที่ปลาย 2 ข้างของกล่องอยู่ที่ที่ค่าควอไทล์ที่ 1 และค่าควอไทล์ที่ 3 ซึ่งกล่องนี้จะบรรจุข้อมูงบริเวณกลาง ๆ การกระจายไว้ 50 %
2. ภายในกล่อง ลากเส้นตรงในแนวดิ่งผ่านค่ามัธยฐาน
3. ลากเส้นในแนวนอนหรือในแนวดิ่งออกไปจากปลายทั้งสองข้างจนถึงค่าต่ำสุด และค่าสูงสุดของข้อมูลชุดนั้น



บอกซ์พลอตจะให้ทราบละเอียดของข้อมูลชุดนั้นน้อยกว่าอิสโตแกรมหรือแผนภาพต้น-ใบ แต่นิยมนำมาใช้เมื่อจะเปรียบเทียบการแจกแจงของข้อมูลหลาย ๆ ชุด ในการสร้างบอกซ์พลอตอาจสร้างรูปในแนวตั้งหรือแนวนอนก็ได้ และควรเขียนกำกับมาตราส่วนไว้ด้วย เมื่อมองที่บอกซ์พลอต จุดแรกที่เราต้องมองหาก็คือ มัธยฐาน ซึ่งเป็นจุดศูนย์กลางของการแจกแจง จากนั้นจุงดูการกระจายโดยค่าควอไทล์จะแสดงถึงการกระจายของครึ่งหนึ่งของข้อมูลบริเวณกลาง ๆ และค่าต่ำสุดกับค่าสูงสุดซึ่งแสดงการกระจายของข้อมูลทั้งหมดในชุดนั้น



จากบอกซ์พลอต จะทำให้เราทราบว่าการแจกแจงของข้อมูลสมมาตรหรือเบ้ ถ้าของมูลมีการแจกแจงที่สมมาตร ค่าควอไทล์ที่ 1 และค่าควอไทล์ที่ 3 จะมีระยะห่างจากมัธยฐานเท่ากัน ในกรณีที่ระยะทางระหว่างค่าควอไทล์ที่ 1 กับค่ามัธยฐานยาวกว่าระยะทางระหว่างค่ามัธยฐานกับค่าควอไทล์ที่ 3 แสดงว่าการแจกแจงของข้อมูลมีลักษณะเบ้ซ้าย

การใช้จำนวน 5 จำนวนสรุปลักษณะของข้อมูลจะดีกว่าการสรุปโดยใช้ค่าเฉลี่ย และส่วนเบี่ยงเบนมาตราฐาน ในกรณีที่การแจกแจงของข้อมูลมีลักษณะเบ้ หรือข้อมูลมีค่าผิดปกติ (outlier) รวมอยู่ด้วย

ตัวอย่างที่ 40 ข้อมูลต่อไปนี้เป็นยอดขายรายปีของร้านขายน้ำ 21 แห่ง หน่วยเป็นล้านบาท ซึ่งเรียงลำดับข้อมูลแล้ว

608	739	1356	1374	1850	1872	2127	2459	2818	3653	
4019	4341	5794	5452	6452	7478	8305	8408	8879	10498	14138

จงสร้างบอกซ์พลอตแสดงยอดขายรายปี และจากบอกซ์พลอตที่สร้างได้สามารถสรุปอะไรได้บ้าง

ควรระลึกไว้เสมอว่ารูปกราฟจะให้ภาพทั้งหมดของการแจกแจงได้ดีที่สุด ตัวเลขที่ใช้บอกจุดศูนย์กลางและการกระจายจะให้ความจริงเฉพาะของการแจกแจง แต่ไม่ได้บอกให้เราทราบเกี่ยวกับรูปร่างทั้งหมด โดยเฉพาะในกรณีที่การแจกแจงมีค่าฐานนิยมหลายค่าหรือข้อมูลมีช่องว่าง (gab) ในบางค่า ซึ่งทำให้เกิดการเข้าใจผิดพลาดได้ ดังนั้น ในการศึกษาลักษณะของข้อมูลจึงควรลงจุดข้อมูล (พลอตกราฟ) เสมอ

ตัวอย่างที่ 41 จากข้อมูลแผนภาพต้น-ใบ จงสร้างบอกซ์พลอต

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

แบบฝึกทักษะ 5

1. จงหา $Q_3 - Q_1$ ถ้ากำหนดความสูงของนักเรียน 15 คน เป็นดังนี้ (หน่วยเป็นเซนติเมตร)
   145 142 160 154 146 142 150 144 152 148 140 152 158 143 145

2. ข้อมูลชุดหนึ่งมี 401 ตัว เมื่อจัดเรียงลำดับจากน้อยไปหามาก พบว่าลำดับของข้อมูลตัวที่ 198 ถึง 202 คือ 87, 88, 92, 95 และ 97 ตามลำดับ จงหา P_{50} ของข้อมูลชุดนี้

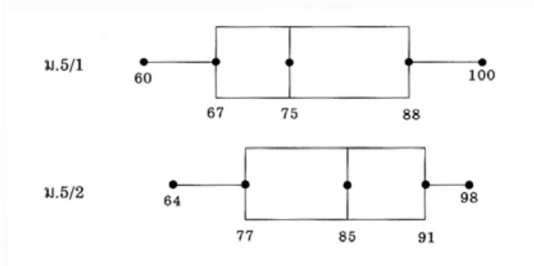
3. จากข้อมูล 18.3 20.6 19.3 22.4 20.2 18.8 19.7 20.0 19.6 18.8 จงหา $P_{89} - Q_3 + D_4$

4. ตารางต่อไปนี้แสดงคะแนนของนักเรียน 150 คน

คะแนน	จำนวนนักเรียน
30 – 39	8
40 – 49	10
50 – 59	12
60 – 69	45
70 – 79	50
80 – 89	20
90 – 99	5

ถ้าแบ่งนักเรียนทั้ง 150 คน ออกเป็น 3 กลุ่ม คือกลุ่มสูง 20% กลุ่มต่ำ 10% และที่เหลือเป็นกลุ่มปกติ
จงหาคะแนนที่ใช้แบ่งนักเรียนทั้ง 3 กลุ่มนี้เป็นเท่าใด

5. จากแผนภาพ แสดงผลสรุปของคะแนนสอบวิชาคณิตศาสตร์ของ ม.5/1 และ ม.5/2



1. จงหาว่านักเรียน ม.5/1 ที่ได้คะแนนอยู่ในกลุ่ม 25% ต่ำสุด มีคะแนนต่ำสุดและคะแนนสูงสุดเป็นเท่าใด
2. นักเรียน ม.5/2 ที่ได้คะแนนมากกว่าหรือเท่ากับ 91 คะแนน มีประมาณกี่เปอร์เซ็นต์
3. มีนักเรียน ม.5/1 กี่เปอร์เซ็นต์ ที่ได้คะแนนน้อยกว่าหรือเท่ากับ 75 คะแนน
4. มีนักเรียน ม.5/2 กี่เปอร์เซ็นต์ ที่ได้คะแนนมากกว่าหรือเท่ากับ 77 คะแนน
5. ถ้าในการสอบครั้งนี้ผู้สอนจะให้ระดับคะแนน 4 แก่ผู้ที่ได้ 80 คะแนนขึ้นไป ห้องใดจะมีผู้ได้ระดับคะแนน 4 มากกว่ากัน เพราะเหตุใด

6. การวัดการกระจายของข้อมูล

Measures of Dispersion

การวัดการกระจาย (Measures of Dispersion) เป็นสถิติประเภทหนึ่งที่คำนวณออกมาเป็นตัวเลข เพื่อใช้อธิบายลักษณะการกระจายของข้อมูล การที่ข้อมูลชุดหนึ่งๆ ประกอบด้วยคะแนนที่มีค่าต่างๆ กันเราเรียกว่า เป็นข้อมูลที่มีการกระจาย ถ้าข้อมูลชุดนั้นประกอบด้วยคะแนนที่มีค่าต่างกันมาก เรียกว่า เป็นข้อมูลที่มีกระจายมาก ถ้าข้อมูลชุดนั้นประกอบด้วยคะแนนที่มีค่าต่างกันน้อย เรียกว่า เป็นข้อมูลที่มีการกระจายน้อย และถ้าข้อมูลชุดนั้นประกอบด้วยคะแนนที่มีค่าเท่ากันหมด เรียกว่า เป็นข้อมูลที่ไม่มีการกระจาย ดังตัวอย่าง

ตารางที่ 1 แสดงข้อมูลซึ่งมีการกระจายต่างกัน

ข้อมูลชุดที่	คะแนนในชุดข้อมูล	ลักษณะการกระจาย
1	7 10 35 70 100	มีการกระจายมาก
2	50 58 60 61 67	มีการกระจายน้อย
3	30 30 30 30 30	ไม่มีการกระจาย

การวัดการกระจายนิยมใช้ควบคู่กับการวัดแนวโน้มเข้าสู่ส่วนกลาง เพราะจะช่วยอธิบายลักษณะของข้อมูลได้ชัดเจนขึ้น ทั้งนี้ เนื่องจากการวัดแนวโน้มเข้าสู่กลางเป็นเพียงการบอกค่ากลางของข้อมูลชุดนั้น แต่เราก็ยังไม่ทราบชัดเจนถึงลักษณะการกระจายของข้อมูลว่าคะแนนต่างๆ ในชุดข้อมูลนั้นมีค่าใกล้เคียงกัน หรือแตกต่างกันมาก ถ้าเรามีทั้งค่าแนวโนม้เข้าสู่ส่วนกลางและค่าการกระจายก็จะทำให้เข้าใจลักษณะข้อมูลนั้นได้ชัดเจนขึ้นมากกว่ามีแต่ค่าแนงโน้มเข้าสู่ส่วนกลางเพียงอย่างเดียว

ตัวอย่างเช่น นักเรียน 2 กลุ่ม ได้รับการทดสอบก่อนเรียนในวิชาเดียวกัน ด้วยข้อสอบชุดเดียวกัน ผลปรากฏว่าคะแนนทดสอบของทั้ง 2 กลุ่มมีค่าเฉลี่ย (Mean) 40 คะแนนเท่ากัน

ถ้าผู้สอนทราบเพียงว่า นักเรียน 2 กลุ่มทำแบบทดสอบก่อนเรียน ได้ค่าเฉลี่ยเท่ากัน คือ 40 คะแนน ก็จะเข้าใจเพียงว่านักเรียน 2 กลุ่มนี้มีความรู้พื้นฐานพอๆ กัน แต่จะไม่ทราบว่าการกระจายของคะแนนหรือความรู้พื้นฐานของนักเรียนแต่ละกลุ่มเป็นอย่างไร นักเรียนแต่ละกลุ่มมีความรู้พื้นฐานแตกต่างกันมากน้อยเพียงไรลองพิจารณาข้อมูลคะแนนในแต่ละกลุ่มดังต่อไปนี้

กลุ่ม ก	45	31	60	54	21	28	41	(Mean = 40)
กลุ่ม ข	39	45	30	41	32	50	43	(Mean = 40)

เมื่อพิจารณาอย่างคร่าว ๆ จะพบว่า ในกลุ่ม ก คะแนนแตกต่างกันมากกว่ากลุ่ม ข นั่นคือ นักเรียนในกลุ่ม ก มีความรู้พื้นฐานแตกต่างกันมากกว่ากลุ่ม ข ช่วยให้ผู้สอนเข้าใจถึงความแตกต่างของความรู้พื้นฐานของนักเรียนในแต่ละกลุ่มว่าต่างกัน กลุ่ม ก นักเรียนมีความรู้พื้นฐานแตกต่างกันมากกว่า ในกลุ่ม ข ซึ่งจะเป็นประโยชน์ในการจัดการเรียนการสอนต่อไป

จะเห็นได้ว่า การทราบเพียงค่าแนวโน้มเข้าสู่ส่วนกลางเพียงอย่างเดียวก็อาจอธิบายลักษณะของข้อมูลได้ไม่สมบูรณ์ ถ้าทราบลักษณะการกระจายด้วยก็จะช่วยให้เข้าใจเกี่ยวกับข้อมูลละเอียดขึ้น และเป็นประโยชน์มาก

การวัดการกระจายของข้อมูล แบ่งได้เป็น 2 วิธี คือ

1. การวัดการกระจายสัมบูรณ์ (Absolute Variation) คือการวัดการกระจายของข้อมูลเพียงชุดเดียว เพื่อดูว่าข้อมูลชุดนั้นแต่ละค่ามีความแตกต่างกันมากหรือน้อยเพียงไร นิยมใช้กันอยู่ 4 ชนิด คือ

- พิสัย (range)
- ส่วนเบี่ยงเบนควอร์ไทล์ (quartile deviation)
- ส่วนเบี่ยงเบนเฉลี่ย (mean deviation หรือ average deviation)
- ส่วนเบี่ยงเบนมาตรฐาน (standard deviation)

2. การวัดการกระจายสัมพัทธ์ (Relative Variation) คือการวัดการกระจายของข้อมูลที่มากกว่า 1 ชุด โดยใช้อัตราส่วนของค่าที่ได้จากการวัดการกระจายสัมบูรณ์ กับค่ากลางของข้อมูลนั้นๆ เพื่อใช้ในการเปรียบเทียบการกระจายของข้อมูลเหล่านั้น มีอยู่ 4 ชนิด คือ

- สัมประสิทธิ์ของพิสัย (coefficient of range)
- สัมประสิทธิ์ของส่วนเบี่ยงเบนควอร์ไทล์ (coefficient of quartile deviation)
- สัมประสิทธิ์ของส่วนเบี่ยงเบนเฉลี่ย (coefficient of average deviation)
- สัมประสิทธิ์ของความแปรผัน (coefficient of variation)

6.1 การวัดการกระจายสัมบูรณ์ (Absolute Variation)

พิสัย (Range)

เป็นค่าที่ใช้วัดการกระจายของข้อมูล โดยการหาความแตกต่างระหว่างค่าสูงสุดกับค่าต่ำสุดของข้อมูลชุดใดชุดหนึ่ง ซึ่งอาจอยู่ในรูปของค่าผลต่างระหว่างค่าสูงสุดกับค่าต่ำสุดหรืออยู่ในรูปช่วงคะแนนจากค่าต่ำสุดถึงค่าสูงสุด นับเป็นวิธีการกระจายอย่างคร่าวๆ และง่ายที่สุด เนื่องจากคะแนนเพียง 2 ค่า เท่านั้นในการคำนวณ คือค่าสูงสุดและค่าต่ำสุด คะแนนค่าอื่นๆ ไม่ได้นำเอามาใช้เลย

ถ้าพิสัยมีค่ามากแสดงว่ามีการกระจายมาก ถ้าพิสัยมีค่าน้อยแสดงว่ามีการกระจายน้อย

การวัดการกระจายด้วยค่าพิสัย มักใช้ควบคู่กับการวัดแนวโน้มเข้าสู่ส่วนกลางด้วยค่าฐานนิยม (Mode) หรืออาจใช้ควบคู่กับการวัดแนวโน้มเข้าสู่ส่วนกลางวิธีอื่นๆ ในกรณีที่มีข้อมูลจำนวนน้อยหรือเมื่อต้องการทราบการกระจายอย่างคร่าวๆ โดยรวดเร็ว

วิธีการหาค่าพิสัย

- กรณีของข้อมูลที่ไม่ได้แจกแจงความถี่

  ใช้สูตร พิสัย = ค่าสูงสุด − ค่าต่ำสุด

  หรือ Range = $x_{\max} - x_{\min}$
- กรณีของข้อมูลที่แจกแจงความถี่โดยแบ่งเป็นอัตราภาคชั้น

  ใช้สูตร พิสัย = ขอบเขตบนของอันตรภาคชั้นที่มีข้อมูลที่มีค่าสูงสุด − ขอบเขตล่างของอันตรภาคชั้นที่มีข้อมูลที่มีค่าต่ำสุด

ตัวอย่างที่ 42 จงหาพิสัยของข้อมูลคะแนนทดสอบความรู้พื้นฐานก่อนเรียนวิชาสิถิตของนักเรียน 2 กลุ่มดังต่อไปนี้

กลุ่ม ก	45	31	60	54	21	28	41
กลุ่ม ข	39	45	30	41	32	50	43

ตัวอย่างที่ 43 จงหาพิสัยของข้อมูลต่อไปนี้

คะแนน	จำนวนนักเรียน
30 − 39	8
40 − 49	10
50 − 59	12
60 − 69	45
70 − 79	50
80 − 89	20
90 − 99	5

ส่วนเบี่ยงเบนควอร์ไทล์

ส่วนเบี่ยงเบนควอไทล์ (Quartile deviation : Q.D.) เป็นค่าที่ใช้วัดการกระจายข้อข้อมูลรอบๆ ค่ามัธยฐาน (Median) ซึ่งมีค่าเท่ากับครึ่งหนึ่งของผลต่างระหว่างควอไทล์ที่ 3 กับควอไทล์ที่ 1 ถ้าส่วนเบี่ยงเบนควอไทล์มีค่ามากแสดงว่ามีการกระจายมากถ้าส่วนเบี่ยงเบนควอไทล์มีค่าน้อยแสดงว่ามีการกระจายน้อย

วิธีการหาค่าส่วนเบี่ยงเบนควอไทล์

ใช้สูตร

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

จากสูตรจะเห็นได้ว่าค่า Q.D. แสดงถึงการกระจายของคะแนนว่าห่างจากมัธยฐาน (Median) ซึ่งเป็นค่าตำแหน่งกึ่งกลางของชุดข้อมูลมากน้อยเพียงไร จึงมักใช้ควบคู่กันกับการวัดแนวโน้มเข้าสู่ส่วนกลางด้วยค่ามัธยฐาน

ตัวอย่างที่ 44 จงหาค่าส่วนเบี่ยงเบนควอไทล์กรณีข้อมูลไม่ได้แจกแจงความถี่

60 31 25 80 77 52 39 45 68 74

ตัวอย่างที่ 45 จงหาค่าส่วนเบี่ยงเบนควอไทล์ของข้อมูลต่อไปนี้

ชั้นคะแนน	ความถี่	บอบเขต	ความถี่สะสม
93 – 97	8		
88 – 92	9		
83 – 87	7		
78 – 82	4		
73 – 77	7		
68 – 72	5		
63 – 67	4		
58 – 62	2		
53 – 57	2		
48 – 52	2		

แบบฝึกทักษะ 6.1

1. กำหนดข้อมูล 9 14 6 8 5 12 8 6 8 11 9 จงหาพิสัย และส่วนเบี่ยงเบนควอไทล์

2. จงหาพิสัยและส่วนเบี่ยงเบนควอไทล์ ของข้อมูลต่อไปนี้
   ตารางต่อไปนี้แสดงคะแนนของนักเรียน

คะแนน	จำนวนนักเรียน
10 – 14	2
15 – 24	3
25 – 34	6
35 – 44	5
45 – 54	4

ส่วนเบี่ยงเบนเฉลี่ย

ส่วนเบี่ยงเบนเฉลี่ย (Mean deviation : M.D.) เป็นค่าที่ใช้วัดการกระจายของข้อมูลรอบๆ ค่าเฉลี่ย (Mean) โดยการหาค่าเฉลี่ยของผลรวมของผลต่างระหว่างคะแนนแต่ละตัวกับค่าเฉลี่ย ถ้าส่วนเบี่ยงเบนเฉลี่ยมีค่ามากแสดงว่ามีการกระจายมาก ถ้าส่วนเบี่ยงเบนเฉลี่ยมีค่าน้อยแสดงว่ามีการกระจายน้อย

วิธีการหาค่าส่วนเบี่ยงเบนเฉลี่ย

ข้อมูลไม่ไดแจกแจงความถี่	ข้อมูลแจกแจงความถี่
$M.D. = \dfrac{\sum_{i=1}^{N}\left\|x_i - \mu\right\|}{N}$	$M.D. = \dfrac{\sum_{i=1}^{k} f_i\left\|x_i - \mu\right\|}{N}$

(ข้อมูลระดับตัวอย่างยังคงใช้สูตรทำนองเดียวกัน)

ตัวอย่างที่ 46 จงหาค่าส่วนเบี่ยงเบนเฉลี่ย ของข้อมูลดังต่อไปนี้ 4 12 7 6 11

ตัวอย่างที่ 47 จงหาค่าส่วนเบี่ยงเบนเฉลี่ยของข้อมูลต่อไปนี้

ชั้นคะแนน	ความถี่ (f_i)	จุดกึ่งกลาง (x_i)	$\left\|x_i - \bar{x}\right\|$	$f_i\left\|x_i - \mu\right\|$
93 – 97	8			
88 – 92	9			
83 – 87	7			
78 – 82	4			
73 – 77	7			
68 – 72	5			
63 – 67	4			
58 – 62	2			
53 – 57	2			
48 – 52	2			

เนื่องจากการหาค่า M.D. ไม่คำนึงถึงเครื่องหมายของผลต่างระหว่างคะแนนแต่ละตัวกับค่าเฉลี่ยแต่เป็นค่าสัมบูรณ์ (Absolute value) ซึ่งไม่เหมาะกับการวิเคราะห์ทางคณิตศาสตร์ต่อไป จึงไม่เป็นที่นิยมใช้และมีผู้คิดวัดการกระจายที่เหมาะสมมากกว่าขึ้นมา คือ วิธีการหาส่วนเบี่ยงเบนมาตรฐานซึ่งจะได้กล่าวถึงในหัวข้อต่อไป

ส่วนเบี่ยงเบนมาตรฐาน

ส่วนเบี่ยงเบนมาตรฐาน (Standard deviation : S.D.) เป็นการวัดการกระจายของคะแนนรอบๆ ค่าเฉลี่ย (Mean) คล้ายๆ กับส่วนเบี่ยงเบนเฉลี่ย แต่แก้ปัญหาค่าสัมบูรณ์โดยใช้วิธียกกำลังสอง ค่าผลต่างระหว่างคะแนนแต่ละตัวกับค่าเฉลี่ย ทำให้เครื่องหมายลบหมดไปเมื่อหาค่าเฉลี่ยของผลรวม

กรณีข้อมูลไม่ได้แจกแจงความถี่ มีสูตรดังนี้

ค่าส่วนเบี่ยงเบนมาตรฐานของประชากร แทนด้วย σ	ค่าส่วนเบี่ยงเบนมาตรฐานของตัวอย่าง แทนด้วย s
$\sigma = \sqrt{\dfrac{\sum_{i=1}^{N}(x_i-\mu)^2}{N}}$	$s = \sqrt{\dfrac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}}$

จะเห็นได้ว่า ค่าส่วนเบี่ยงเบนมาตรฐานก็คือรากที่สองของความแปรปรวน เขียนเป็นสูตรได้

ค่าส่วนเบี่ยงเบนมาตรฐาน^2 = ความแปรปรวน

ดั้งนั้นถ้าทราบความแปรปรวนของข้อมูลแล้ว จะสามารถหาค่าส่วนเบี่ยงเบนมาตรฐานได้ โดยการถอดรากที่สองของความแปรปรวนนั้น เขียนเป็นสูตรในรูปสัญลักษณ์ได้ดังนี้

$$\sigma = \sqrt{\sigma^2} \qquad \text{หรือ} \qquad s = \sqrt{s^2}$$

ในทางกลับกันถ้าทราบค่าส่วนเบี่ยงเบนมาจรฐานแล้ว จะสามารถหาค่าความแปรปรวนได้ โดยการยกกำลังสองค่าส่วนเบี่ยงเบนมาตรฐานนั้น

ส่วนสูตรในการคำนวณหาค่าความแปรปรวนและค่าส่วนเบี่ยงเบนมาตรฐาน ซึ่งเป็นสูตรที่อยู่ในรูปคะแนนเบี่ยงเบน (ผลต่าง) ตามนิยามที่กล่าวมาข้างต้นเขียนสรุปได้ดังต่อไปนี้

สำหรับประชากร

	ความแปรปรวน	**ส่วนเบี่ยงเบนมาตรฐาน**
กรณีข้อมูลไม่ได้แจกแจงความถี่		
• สูตรในรูปคะแนนเบี่ยงเบน	$\sigma^2 = \frac{\Sigma(x-\mu)^2}{N}$	$\sigma = \sqrt{\frac{\Sigma(x-\mu)^2}{N}}$
• สูตรที่อยู่ในรูปคะแนนดิบ	$\sigma^2 = \frac{\Sigma x^2}{N} - \mu^2$	$\sigma = \sqrt{\frac{\Sigma x^2}{N} - \mu^2}$
กรณีข้อมูลแจกแจงความถี่		
• สูตรในรูปคะแนนเบี่ยงเบน	$\sigma^2 = \frac{\Sigma f(x-\mu)^2}{N}$	$\sigma = \sqrt{\frac{\Sigma f(x-\mu)^2}{N}}$
• สูตรที่อยู่ในรูปคะแนนดิบ	$\sigma^2 = \frac{\Sigma fx^2}{N} - \mu^2$	$\sigma = \sqrt{\frac{\Sigma fx^2}{N} - \mu^2}$

สำหรับตัวอย่าง

	ความแปรปรวน	**ส่วนเบี่ยงเบนมาตรฐาน**
กรณีข้อมูลไม่ได้แจกแจงความถี่		
• สูตรในรูปคะแนนเบี่ยงเบน	$s^2 = \frac{\Sigma(x-\bar{x})^2}{n-1}$	$s = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}}$
• สูตรที่อยู่ในรูปคะแนนดิบ	$s^2 = \frac{\sum x^2 - n\bar{x}^2}{n-1}$	$s = \sqrt{\frac{\sum x^2 - n\bar{x}^2}{n-1}}$
กรณีข้อมูลแจกแจงความถี่		
• สูตรในรูปคะแนนเบี่ยงเบน	$s^2 = \frac{\Sigma f(x-\bar{x})^2}{n-1}$	$s = \sqrt{\frac{\Sigma f(x-\bar{x})^2}{n-1}}$
• สูตรที่อยู่ในรูปคะแนนดิบ	$s^2 = \frac{\Sigma fx^2 - n\bar{x}^2}{n-1}$	$s = \sqrt{\frac{\Sigma fx^2 - n\bar{x}^2}{n-1}}$

ตัวอย่างที่ 48 จงหาค่าความแปรปรวนและส่วนเบี่ยงเบนมาตรฐานของคะแนนผลสัมฤทธิ์ทางการเรียนวิชาสถิติเพื่อการวิจัยของนักเรียนกลุ่มตัวอย่างจำนวน 10 คน ซึ่งมีค่าดังต่อไปนี้

87 61 75 77 85 92 83 73 65 58

วิธีทำ ก. สูตรในรูปคะแนนเบี่ยงเบน

วิธีทำ ข. สูตรในรูปคะแนนดิบ

ตัวอย่างที่ 49 ครูวิทยาศาสตร์ท่านหนึ่ง ต้องการทราบความสามารถในการใช้อุปกรณ์การทดลอง วิทยาศาสตร์ของนักเรียนห้องหนึ่ง จึงสุ่มตัวอย่างนักเรียนมา 25 คน และวัดความสามารถในการทดลองวิทยาศาสตร์ของนักเรียนแต่ละคน คะแนนที่ได้นำมาแจกแจงความถี่ได้ดังต่อไปนี้ จงคำนวณหาความแปรปรวนและส่วนเบี่ยงเบนมาตรฐานของคะแนนความสามารถในการใช้อุปกรณ์การทดลองวิทยาศาสตร์ของนักเรียนห้องนี้

ชั้นคะแนน	**ความถี่**
79 – 81	2
76 – 78	3
73 – 75	4
70 – 72	7
67 – 69	5
64 – 66	2
61 – 63	1
รวม	**25**

แบบฝึกทักษะ 6.2

1. กำหนดข้อมูลประชากร 9 14 6 8 5 12 8 6 8 11 9 จงหาส่วนเบี่ยงเบนเฉลี่ย และส่วนเบี่ยงเบนมาตรฐาน

2. จงหาส่วนเบี่ยงเบนเฉลี่ย และส่วนเบี่ยงเบนมาตรฐาน ของข้อมูลประชากรต่อไปนี้
   ตารางต่อไปนี้แสดงคะแนนของนักเรียน

คะแนน	**จำนวนนักเรียน**
10 – 14	2
15 – 24	3
25 – 34	6
35 – 44	5
45 – 54	4

จะเห็นได้ว่าการหาส่วนเบี่ยงเบนมาตรฐานนั้นมีการคำนวณที่ยุ่งยาก และหลายขั้นตอน นอกจากสูตรดังกล่าวแล้ว เราสามารถหาส่วนเบี่ยงเบนมาตรฐานสำหรับประชากรโดยวิธีทอนค่าของข้อมูล ดังนี้

$$\sigma = I\sqrt{\frac{\sum f(d-\bar{d})^2}{N}} \qquad \text{หรือ} \qquad \sigma = I\sqrt{\frac{\sum fd^2}{N} - \bar{d}^2}$$

ตัวอย่างที่ 50 จงหาส่วนเบี่ยงเบนมาตรฐานโดยวิธีทอนค่าของข้อมูลประชากรต่อไปนี้

ชั้นคะแนน	ความถี่
79 – 81	2
76 – 78	3
73 – 75	4
70 – 72	7
67 – 69	5
64 – 66	2
61 – 63	1
รวม	**25**

ตัวอย่างที่ 51 จงหาสูตรเบี่ยงเบนมาตรฐานโดยวิธีทอนค่าสำหรับข้อมูลระดับตัวอย่าง

โจทย์ท้าทาย

การหาส่วนเบี่ยงเบนมาตรฐานและความแปรปรวนเมื่อกำหนดค่ากลางอื่นๆ มาให้

ตัวอย่างที่ 52 ข้อมูลประชากรชุดหนึ่ง ถ้า $\sum_{i=1}^{10} x_i = 60$ และ $\sum_{i=1}^{10} (x_i - 5)^2 = 46$ จงหาส่วนเบี่ยงเบนมาตรฐาน

ตัวอย่างที่ 53 ข้อมูลชุดหนึ่งมี N ตัว มี $\sqrt{\frac{\sum_{i=1}^{N}(x_i-30)^2}{N}} = 30$ และค่าเฉลี่ยเลขคณิตเท่ากับ 33 จงหาความแปรปรวนของข้อมูลชุดนี้

การหาส่วนเบี่ยงเบนมาตรฐานและความแปรปรวนเมื่อมีการอ่านข้อมูลผิดพลาด

ตัวอย่างที่ 54 ข้อมูลประชากรชุดหนึ่งมี 100 ตัว หาค่าเฉลี่ยเลขคณิตได้ 9 และส่วนเบี่ยงเบนมาตรฐานได้ 5 แต่เมื่อนำมาทบทวนภาพหลังพบว่าสิ่งที่คำนวณนั้นผิด เพราะผู้ทำการคำนวณอ่านข้อมูลผิดไป 1 ตัว คือ จาก 1.0 อ่านเป็น 10 ดังนั้นค่าเฉลี่ยเลขคณิคและส่วนเบี่ยงเบนมาตรฐานที่ถูกต้องเป็นเท่าใด

สมบัติที่สำคัญของส่วนเบี่ยงเบนมาตรฐาน

- ส่วนเบี่ยงเบนมาตรฐานมีค่าเป็นบวกเสมอ
- ถ้าคำนวณหาส่วนเบี่ยงเบนมาตรฐานโดยใช้ค่ากลางของมูลชนิดชนิดอื่นๆ ที่ไม่ใช่ค่าเฉลี่ยเลขคณิต ส่วนเบี่ยงเบนมาตรฐานที่จะได้จะมีค่ามากกว่าส่วนเบี่ยงเบนมาตรฐานที่ใช้ค่าเฉลี่ยเลขคณิตเสมอ นั่นคือ

  $$\sqrt{\frac{\Sigma f(x-a)^2}{N}} > \sqrt{\frac{\Sigma f(x-\mu)^2}{N}}$$

  อสมการนี้เป็นจริงเสมอ เมื่อ a เป็นจำนวนจริงใดๆ ที่ไม่เท่ากับค่าเฉลี่ยเลขคณิต
- ถ้ามีข้อมูล 2 ชุด ประกอบด้วยข้อมูล N_1 และ N_2 จำนวน มีค่าเฉลี่ยเลขคณิตเท่ากัน แต่มีความแปรปรวนเป็น σ_1^2 และ σ_2^2 สำหรับข้อมูลชุดที่ 1 และ 2 ตามลำดับ ความแปรปรวนของข้อมูลทั้งสองชุดจะเท่ากับ $\frac{N_1\sigma_1^2+N_2\sigma_2^2}{N_1+N_2}$

ตัวอย่างที่ 55 ข้อมูล 2 ชุด ชุดแรกมี 5 จำนวน ความแปรปรวน 18 ชุดหลังมี 3 จำนวน ความแปรปรวน 24 ข้อมูลทั้งสองชุดมีค่าเฉลี่ยเลขคณิตเท่ากัน จงหาความแปรปรวนรวมของข้อมูลทั้งสองชุด

ข้อสังเกตเกี่ยวกับใช้ค่าการวัดการกระจาย

ค่าพิสัย

1. พิสัยเป็นการวัดการกระจายโดยใช้คะแนนเพียง 2 ตัว ไม่ได้นำคะแนนทุกตัวมาใช้ในการคำนวณจึงเป็นวิธีการกระจายอย่างหยาบๆ

2. พิสัยเหมาะสำหรับวัดการกระจายอย่างคร่าวๆ เมื่อต้องการทราบค่าการกระจายอย่างรวดเร็ว เพราะใช้เวลาน้อยในการคำนวณ

3. พิสัยเหมาะกับชุดข้อมูลขนาดเล็กมากกว่าขนาดใหญ่ถ้าข้อมูลใหญ่มีแนวโน้มค่าพิสัยสูง

4. ไม่ควรใช้พิสัยในการเปรียบเทียบการกระจายของข้อมูลที่มีขนาดไม่เท่ากัน ถ้าเป็นข้อมูลกลุ่มใหญ่มีแนวโน้มที่ค่าพิสัยจะสูง

ค่าส่วนเบี่ยงเบนควอไทล์

1. ส่วนเบี่ยงเบนควอไทล์ เป็นการวัดการกระจายที่ดีกว่าการวัดด้วยค่าพิสัย แต่ก็ยังใช้เพียงบางค่าไม่ได้ใช้ข้อมูลทุกค่าในการคำนวณ

2. ส่วนเบี่ยงเบนควอไทล์ เหมาะสำหรับใช้วัดการกระจายกรณีมีคะแนนบางค่าสูงหรือต่ำกว่าคะแนนตัวอื่นๆ ในชุดมาก

3. ส่วนเบี่ยงเบนควอไทล์เหมาะสำหรับใช้ควบคู่กับการวัดแนวโน้มเข้าสู่ส่วนกลางด้วยค่ามัธยฐาน

ค่าส่วนเบี่ยงเบนเฉลี่ย

1. ส่วนเบี่ยงเบนเฉลี่ย เป็นการวัดการกระจายที่ละเอียดกว่าการวัดด้วยค่าพิสัยและส่วนเบี่ยงเบน ควอไทล์ เพราะได้ใช้คะแนนทุกๆ ตัวในการคำนวณ

2. การคำนวณค่าส่วนเบี่ยงเบนเฉลี่ยไม่ได้คำนึงถึงเครื่องหมายของผลต่างระหว่างคะแนนแต่ละตัวกับค่าเฉลี่ย ซึ่งขัดต่อหลักคณิตศาสตร์ จึงไม่เป็นที่นิยมใช้

ค่าส่วนเบี่ยงเบนมาตรฐาน

1. ส่วนเบี่ยงเบนมาตรฐาน เป็นการวัดการกระจายของข้อมูลที่ใช้คะแนนทุกตัว ในการคำนวณจึงเป็นการวัดการกระจายที่ละเอียดกว่าการหาโดยพิสัย พิสัยควอไทล์ และส่วนเบี่ยงเบนควอไทล์

2. ส่วนเบี่ยงเบนมาตรฐาน เป็นการวัดการกระจายของข้อมูลที่นิยมใช้มากที่สุดโดยเฉพาะอย่างยิ่งในการวิจัย

3. ค่าส่วนเบี่ยงเบนมาตรฐาน เป็นค่าการกระจายที่มีหน่วยเช่นเดียวกับหน่วยของข้อมูลที่เก็บรวบรวมมา ส่วนค่าความแปรปรวน ซึ่งมีค่าเป็นกำลังสองของส่วนเบี่ยงเบนมาตรฐานเป็นค่าการกระจายที่มีหน่วยเป็นกำลังสองของหน่วยของข้อมูลที่เก็บรวบรวมมา

ท้าทาย

1. จงแสดงว่า ถ้าข้อมูลมีค่าเท่ากันหมดทุกตัว ส่วนเบี่ยงเบนมาตรฐานจะมีค่าเท่ากับศูนย์ หรือถ้าส่วนเบี่ยงเบนมาตรฐานมีค่าเท่ากับศูนย์ แล้วข้อมูลแต่ละตัวจะมีค่าเท่ากับค่าเฉลี่ยเลขคณิต

2. จงแสดงว่า ถ้านำจำนวนจริงไปบวกหรือลบกับข้อมูลแต่ละตัว ส่วนเบี่ยงมาตรฐานของข้อมูลชุดใหม่ จะมีค่าเท่าเดิม

3. จงแสดงว่า ถ้านำจำนวนจริงไปคูณหรือหารกับข้อมูลแต่ละตัว ส่วนเบี่ยงมาตรฐานของข้อมูลชุดใหม่ จะมีค่าเท่ากับ |a| เท่าของส่วนเบี่ยงมาตรฐานของข้อมูลชุดเดิม

4. ถ้าความสัมพันธ์ระหว่าง x_i และ y_i เป็นสมการเชิงเส้น $Y = aX + b$ โดยที่ a และ b เป็นค่าคงตัว จะได้ $\sigma_y = |a|\sigma_x$

5. ข้อมูล 2 ชุดมีจำนวนเท่ากัน เมื่อทำการคำนวณหาค่าเฉลี่ยเลขคณิตปรากฏว่า $\mu_1 : \mu_2 = 3 : 5$ และได้ส่วนเบี่ยงเบนมาตรฐานเท่ากัน ถ้าข้อมูลชุดหนึ่งคือ 1, 4, 6, 9, 10 จงหาข้อมูลอีกชุดหนึ่ง

6.2 การวัดการกระจายสัมพัทธ์

ในกรณีที่ต้องการเปรียบเทียบการกระจายของข้อมูลต่างกลุ่ม (ต่างชุด) ถ้าข้อมูลแต่ละชุดเป็นคะแนนที่มีหน่วยวัดเดียวกัน คะแนนเต็มเท่ากัน ขนาดเท่ากัน และค่าแนวโน้มเข้าสู่ส่วนกลางเท่ากันก็สามารถนำค่าการกระจายมาเปรียบเทียบกันได้เลย

ตัวอย่างเช่น คะแนนสอบวิชาภาษาไทยของนักเรียน 2 ห้อง ซึ่งมีจำนวนนักเรียนเท่ากัน สอบด้วยข้อสอบชุดเดียวกัน มีคะแนนเฉลี่ยและส่วนเบี่ยงเบนมาตรฐาน ดังนี้

ห้อง	**ค่าเฉลี่ย**	**ส่วนเบี่ยงเบนมาตรฐาน**
1	60	10
2	60	12

กรณีนี้สามารถบอกได้ว่า คะแนนสอบวิชาภาษาไทยของนักเรียนห้อง 2 มีการกระจายมากกว่าห้อง 1

แต่ถ้าเป็นกรณีที่ข้อมูลแต่ละชุด เป็นคะแนนที่มีหน่วยวัดต่างกันหรือมีคะแนนเต็มไม่เท่ากันหรือขนาดไม่เท่ากันหรือค่าแนวโน้มเข้าสู่ส่วนกลางไม่เท่ากัน จะไม่สามารถนำค่าการกระจายมาเปรียบเทียบกันได้ทันทีแต่ต้องคำนวณหาค่าสัมประสิทธิ์การกระจาย (Coefficient of Dispersion) ของคะแนนแต่ละชุดแล้วจึงนำค่าสัมประสิทธิ์การกระจายนั้นมาเปรียบเทียบกัน ตัวอย่างเช่น คะแนนสอบวิชาภาษาไทยกับคณิตศาสตร์ของนักเรียนห้องหนึ่งมีค่าดังนี้

วิชา	**คะแนนเต็ม**	**ค่าเฉลี่ย**	**ส่วนเบี่ยงเบนมาตรฐาน**
ภาษาไทย	100	60	10
คณิตศาสตร์	150	90	12

กรณีนี้ไม่สามารถบอกได้ว่าคะแนนสอบวิชาใดมีการกระจายมากกว่ากัน แม้ว่าส่วนเบี่ยงเบนมาตรฐานของคะแนนวิชาคณิตศาสตร์จะมีค่ามากกว่าก็ตาม กรณีนี้จะต้องหาค่าสัมประสิทธิ์การกระจายเพื่อนำมาเปรียบเทียบกัน

การหาค่าสัมประสิทธิ์การกระจายมีหลายชนิด แล้วแต่ชนิดของสถิติที่ใช้วัดการกระจายดังต่อไปนี้

- **กรณีวัดการกระจายด้วยพิสัย**

  ใช้สัมประสิทธิ์ของพิสัย (Coefficient of Range : C.R.)

$$\text{C.R.} = \frac{x_{\max} - x_{\min}}{x_{\max} + x_{\min}}$$

- **กรณีวัดการกระจายส่วนเบี่ยงเบนควอไทล์**

  ใช้สัมประสิทธิ์ของส่วนเบี่ยงเบนควอไทล์ (Coefficient of Quartile Deviation : C.Q.)

$$\text{C.Q.} = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

- **กรณีวัดการกระจายด้วยส่วนเบี่ยงเบนเฉลี่ย**

  ใช้สัมประสิทธิ์ของส่วนเบี่ยงเบนเฉลี่ย (coefficient of average deviation)

$$\text{C.A.} = \frac{M.D.}{\mu} \text{ หรือ } \frac{M.D.}{\overline{x}}$$

- **กรณีวัดการกระจายด้วยส่วนเบี่ยงเบนมาตรฐาน**

  ใช้สัมประสิทธิ์ของความแปรผัน (Coefficient of Variation : C.V.)

$$\text{C.V.} = \frac{\sigma}{\mu} \text{ หรือ } \frac{s}{\overline{x}}$$

ตัวอย่างที่ 56 นักศึกษาปริญญาโท 2 กลุ่ม มีอายุดังนี้

กลุ่มที่ 1	30	34	38	28	35	27	42	35	37	39
กลุ่มที่ 2	28	35	24	33	44	26	33	37	40	29

จงเปรียบเทียบการกระจายของอายุของนักศึกษา 2 กลุ่มนี้ เมื่อวัดค่าการกระจายด้วยพิสัย

วิธีทำ ใช้สูตร สัมประสิทธิ์ของพิสัย C.R. $= \dfrac{x_{\max} - x_{\min}}{x_{\max} + x_{\min}}$

$$C.R._1 = \frac{42-27}{42+27} = \frac{15}{69} = 0.217$$

$$C.R._2 = \frac{44-24}{44+24} = \frac{20}{68} = 0.294$$

$$C.R._2 > C.R._1$$

$\therefore$ อายุของนักศึกษาปริญญาโทกลุ่ม 2 มีการกระจายมากกว่ากลุ่ม 1

ตัวอย่างที่ 57 จงเปรียบเทียบการกระจายของข้อมูล 2 ชุดนี้

	มัธยฐาน	Q_3	Q_1
ข้อมูลชุดที่ 1	25	40	15
ข้อมูลชุดที่ 2	30	50	10

ตัวอย่างที่ 58 นำวัตถุ 2 ชนิดไปชั่งบนเครื่องชั่ง 5 อัน ได้ผลดังตาราง

วัตถุชนิดที่ 1 (กรัม)	6	7	9	8	12
วัตถุชนิดที่ 2 (กรัม)	50	52	49	55	44

จงหาสัมประสิทธิ์ส่วนเบี่ยงเบนเฉลี่ยของวัตถุทั้งสองชนิด

ตัวอย่างที่ 59 จงเปรียบเทียบการกระจายของข้อมูลคะแนนสอบ 3 วิชา ของนักเรียนกลุ่มหนึ่งซึ่งมีค่าดังนี้

วิชา	**คะแนนเต็ม**	**Mean**	**S.D.**
ภาษาไทย	100	60	10
คณิตศาสตร์	150	90	12
วิทยาศาสตร์	200	110	16

แบบฝึกทักษะ 6.3

1. ถ้าคะแนนสอบวิชาคณิตศาสตร์และวิชาสถิติของนักเรียนเป็นดังตาราง

คะแนนวิชาสถิติ	6	5	4	2	1
คะแนนวิชาคณิตศาสตร์	9	6	5	3	2

จงหาอัตราส่วนของสัมประสิทธ์ของการแปรผันระหว่างคะแนนวิชาสถิติ และคะแนนวิชาคณิตศาสตร์

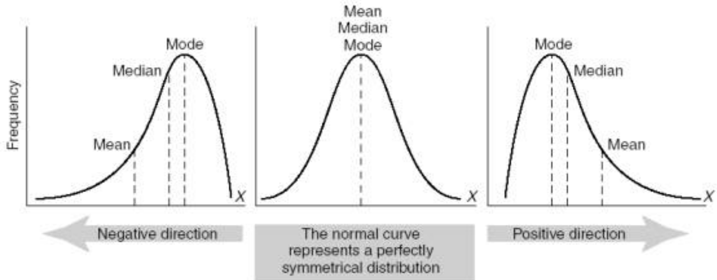
2. บริษัทผลิตหลอดไฟฟ้าแห่งหนึ่ง ผลิตหลอดไฟออกจำหน่าย 2 ชนิด ชนิดแรกอายุการใช้งานเฉลี่ย 1,495 ชั่วโมง ส่วนเบี่ยงเบนมาตรฐาน 280 ชั่วโมง ชนิดที่สอง อายุการใช้งานเฉลี่ย 1,875 ชั่วโมง ส่วนเบี่ยงเบนมาตรฐาน 310 ชั่วโมง จงพิจารณาว่าหลอดไฟชนิดใดมีการกระจายมากกว่ากัน และหลอดไฟใดคุณภาพดีกว่ากัน

ความสัมพันธ์ระหว่างการแจกแจงความถี่ ค่ากลาง และการกระจายของข้อมูล

จากข้อมูลที่มีการแจกแจงความถี่ ถ้านำข้อมูลเหล่านี้มาเขียนให้เป็นเส้นโค้งของความถี่ จะได้เส้นโค้งของความถี่ 3 ลักษณะ (ในระดับสูงขึ้น นักเรียนจะได้ศึกษาเส้นโค้งของความถี่มากกว่า 3 ลักษณะ) ดังนี้

1. เส้นโค้งปกติ หรือเส้นโค้งรูประฆังคว่ำ (normal curve or bell-shaped curve)
2. เส้นโค้งเบ้ลาดทางขวา หรือเส้นโค้งเบ้ทางบวก (positively curve)
3. เส้นโค้งเบ้ลาดทางซ้าย หรือเส้นโค้งเบ้ทางลบ (negatively curve)

ลักษณะของโค้งเป็นดังนี้



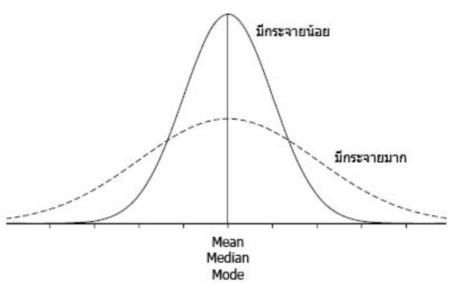
เส้นโค้งของความถี่ของข้อมูลมีความสัมพันธ์กับค่ากลางของข้อมูล คือ

1. โค้งปกติ จะพบว่า ค่าเฉลี่ยเลขคณิต = มัธยฐาน = ฐานนิยม
2. เส้นโค้งเบ้ลาดทางขวา จะพบว่า ฐานนิยม < มัธยฐาน < ค่าเฉลี่ยเลขคณิต
3. เส้นโค้งเบ้ลาดทางซ้าย จะพบว่า ค่าเฉลี่ยเลขคณิต < มัธยฐาน < ฐานนิยม

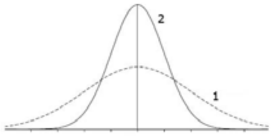
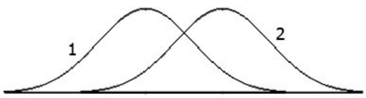
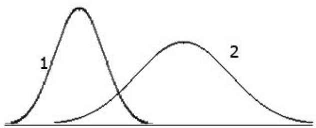
เส้นโค้งของความถี่ที่พบเสมอๆ ไม่ว่าจะเป็นข้อมูลทางด้านประชากร เกษตร สังคม เศรษฐกิจ หรือวิทยาศาสตร์ ส่วนใหญ่มักเป็นข้อมูลที่เกิดขึ้นหรือเป็นไปตามธรรมชาติ และจะมีเส้นโค้งความถี่เป็นรูปเส้นโค้งปกติ เช่น ข้อมูลเกี่ยวกับ ความสูง น้ำหนัก ราคา ผลผลิตทางการเกษตร มักมีรูปเป็นเส้นโค้งปกติ

ลักษณะของเส้นโค้งปกติ

เส้นโค้งปกติมีความโด่งมากหรือน้อยขึ้นอยู่กับการกระจายของข้อมูล ถ้าข้อมูลมีการกระจายมากเส้นโค้งปกติจะโด่งน้อย หรือค่อนข้างแบน แต่ถ้าข้อมูลมีการกระจายน้อย เส้นโค้งปกติจะโด่งมากหรือค่อนข้างสูง ดังรูป



ลักษณะของเส้นโค้งปกติกับการกระจายของข้อมูล

ลักษณะของเส้นโค้งปกติ	บทสรุป
 	• $\mu_1 = \mu_2$ • $\sigma_1 < \sigma_2$ • ข้อมูลชุดที่ 1 กระจายมากกว่าชุด 2
 	• $\mu_1 < \mu_2$ • $\sigma_1 = \sigma_2$ • ข้อมูลชุดที่ 1 กระจายมากกว่าชุด 2 $\frac{\sigma_1}{\mu_1} > \frac{\sigma_2}{\mu_2}$
 	• $\mu_1 < \mu_2$ • $\sigma_1 < \sigma_2$ • ยังสรุปไม่ได้ จนกว่าจะทราบ σ และ μ ของข้อมูลทั้งสองชุด

ตัวอย่างที่ 60 ข้อมูล 2 ชุดมีการแจกแจงความถี่เป็นเส้นโค้งปกติดังรูป

ข้อใดต่อไปนี้กล่าวถูกต้อง

ก. $\mu_1 < \mu_2$ และ $\sigma_1 < \sigma_2$

ข. $\mu_1 > \mu_2$ และ $\sigma_1 < \sigma_2$

ค. $\mu_1 > \mu_2$ และ $\sigma_1 > \sigma_2$

ง. $\mu_1 < \mu_2$ และ $\sigma_1 > \sigma_2$

แบบฝึกทักษะ 6.4

1. จงพิจารณาดูว่า ข้อความต่อไปนี้ ถูกหรือผิด

..................1) พิสัยของข้อมูลใดๆ จะต้องมีค่าเป็นบวกเสมอ

..................2) สัมประสิทธิ์ของพิสัยของข้อมูลอาจเป็นจำนวนลบได้

..................3) ค่าเบี่ยงเบนมาตรฐานและความแปรปรวนของข้อมูลชุดเดียวกันต้องมีค่าต่างกัน

..................4) ค่าเบี่ยงเบนเฉลี่ยและค่าเบี่ยงเบนมาตรฐานของข้อมูลชุดเดียวกัน อาจมีค่าเท่ากันได้

..................5) ถ้าในข้อมูลชุดหนึ่ง มีค่าของข้อมูลทุกตัวเท่ากัน พิสัย ค่าเบี่ยงเบนควอร์ไทล์ ค่าเบี่ยงเบนเฉลี่ย ค่าค่าเบี่ยงเบนมาตรฐาน จะเท่ากันหมด

..................6) ถ้าข้อมูลชุดที่ 1 มีค่าเบี่ยงเบนมาตรฐานมากกว่าข้อมูลชุดที่ 2 แสดงว่าข้อมูลชุดที่ 1 มีการกระจายมากกว่าข้อมูลชุดที่ 2

..................7) ถ้าข้อมูลชุดที่ 1 มีค่าเบี่ยงเบนเฉลี่ยน้อยกว่าข้อมูลชุดที่ 2 แสดงว่าข้อมูลชุดที่ 1 มีการกระจายน้อยกว่าข้อมูลชุดที่ 2

..................8) ค่าเบี่ยงเบนมาตรฐาน และสัมประสิทธิ์ของค่าเบี่ยงเบนเฉลี่ยของข้อมุลชุดเดียวกันจะไม่เท่ากัน

..................9) ค่าเบี่ยงเบนเฉลี่ย และสัมประสิทธิ์ของค่าเบี่ยงเบนเฉลี่ยของข้อมุลชุดเดียวกันจะไม่เท่ากัน

..................10) สัมประสิทธิ์ของความแปรผันของข้อมูลชุดหนึ่ง จะมีค่ามากกว่าสัมประสิทธิ์ของค่าเบี่ยงเบนเฉลี่ยของข้อมูลชุดนั้นเสมอ

..................11) ความแปรปรวนของข้อมูลชุดหนึ่งจะมากกว่าหรือเท่ากับค่าเบี่ยงเบนมาตรฐานของข้อมุลชุดนั้น

..................12) ถ้าความแปรปรวนของข้อมูลชุดหนึ่งมีค่า 0 แสดงว่าค่าของข้อมูลทุกค่าจะเท่ากันหมด

..................13) ข้อมูล 2 ชุดมีค่าเฉลี่ยเลขคณิตต่างกัน จะต้องมีค่าเบี่ยงเบนมาตรฐานต่างกันด้วย

..................14) ข้อมูล 2 ชุดมีค่าเฉลี่ยเลขคณิตต่างกัน จะต้องมีค่าเบี่ยงเบนเฉลี่ยต่างกันด้วย

..................15) ข้อมูล 2 ชุดมีค่าเฉลี่ยเลขคณิตต่างกัน จะต้องมีพิสัยต่างกันด้วย

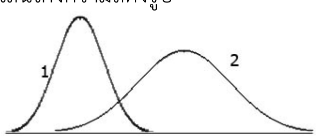
..................16) ข้อมูล 2 ชุดมีค่าเฉลี่ยเลขคณิตเท่ากัน การแจกแจงของข้อมูลจะต้องเหมือนกัน

..................17) ถ้าพิสัยของข้อมูลเท่ากับ 0 ค่าเบี่ยงเบนมาตรฐาน และค่าเบี่ยงเบนเฉลี่ยจะต้องเท่ากัน 0 ด้วย

..................18) ข้อมูล 2 ชุดมีค่าเฉลี่ยเลขคณิตเท่ากัน เส้นโค้งความถี่จะต้องโด่งเท่ากัน

..................19) สัมประสิทธิ์ของพิสัยของข้อมูลชุดหนึ่งเท่ากับ 1 ค่าเฉลี่ยเลขคณิตจะต้องมากกว่า 0

..................20) กำหนดเส้นโค้งความถี่ดังรูป



ค่าเบี่ยงเบนมาตรฐานของข้อมูลชุด 1 มากกว่าข้อมูลชุด 2

..................21) จากข้อ 20 ค่าเบี่ยงเบนเฉลี่ยของข้อชุด 1 น้อยกว่าข้อมูลชุด 2

..................22) ถ้าค่ามากที่สุดของข้อมูล มีค่าเท่ากับค่าเฉลี่ยเลขคณิต ค่าเบี่ยงเบนมาตรฐานจะเท่ากับ 0

..................23) ถ้าค่าต่ำที่สุดของข้อมูล มีค่าเท่ากับค่าเฉลี่ยเลขคณิต ค่าเบี่ยงเบนมาตรฐานจะเท่ากับ 0

..................24) ถ้าสัมประสิทธิ์ของส่วนเบี่ยงเบนควอร์ไทล์เท่ากับ 0 แล้วค่าทุกค่าจะเท่ากันหมด

..................25) ถ้าสัมประสิทธิ์ของพิสัยของข้อมูลมีค่า 0 แล้วค่าทุกค่าจะเท่ากันหมด

2. โรงงานแห่งหนึ่งจ่ายเงินเดือนให้คนงานทั้งหมดมีค่าเฉลี่ยเท่ากับ 4,600 บาท เงินเดือนเฉลี่ยของพนักงานชายและพนักงานหญิงเท่ากับ 5,200 บาท และ 4,200 บาท ถ้าคนงานชายมี 60 % คนงานหญิงจะมีกี่เปอร์เซ็นต์

3. ในการสอบวิชาสถิติ นักเรียนห้อง ก. จำนวน 20 คน สอบได้ค่าเฉลี่ยเลขคณิตเท่ากับ 75 คะแนน นักเรียนห้อง ข. จำนวน 30 คน สอบได้ค่าเฉลี่ยเลขคณิตเท่ากับ 60 คะแนน นักเรียนห้อง ค. จำนวน 25 คน สอบได้ค่าเฉลี่ยเลขคณิตเท่ากับ 65 คะแนน จงหาค่าเฉลี่ยเลขคณิตของวิชาสถิติของนักเรียนทั้ง 3 ห้อง

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

5. ถ้านาย ก. จะต้องตัดสินใจเลือกซื้อหุ้นบริษัทใดบริษัทหนึ่ง จากที่มีให้เลือก 3 บริษัทที่มีอัตราเงินปันผลดังนี้

   บริษัท A เงินปันผลเฉลี่ย 15.6 ต่อปี และค่าส่วนเบี่ยงเบนมาตรฐาน 3.7

   บริษัท B เงินปันผลเฉลี่ย 13.7 ต่อปี และค่าส่วนเบี่ยงเบนมาตรฐาน 2.5

   บริษัท C เงินปันผลเฉลี่ย 18.9 ต่อปี และค่าส่วนเบี่ยงเบนมาตรฐาน 5.8

   ถ้าท่านเป็นนาย ก. ท่านจะตัดสินใจเลือกลงทุนซื้อหุ้นของบริษัทใด