

# สถิติ (Statistics)

สถิติ (Statistics) หมายถึง ตัวเลขที่บอกถึงข้อเท็จจริงต่างๆ หรือข้อมูลที่มีการเปลี่ยนแปลง และเป็นตัวเลขที่อยู่ในลักษณะสรุปรวบยอด ซึ่งประมวลมาได้จากข้อมูลเบื้องต้น โดยการวิเคราะห์ มีระเบียบวิธีการทางสถิติ 4 ขั้นตอน คือ

1. การเก็บรวบรวมข้อมูล (Collection of Data)
2. การนำเสนอข้อมูล (Presentation of Data)
3. การวิเคราะห์ข้อมูล (Analysis of Data)
4. การตีความหมายข้อมูล (Interpretation of Data)

**ข้อมูล (Data)** หมายถึง ข้อเท็จจริงเรื่องใดเรื่องหนึ่งที่ต้องการศึกษา ซึ่งอาจเป็นตัวเลขหรือข้อความ ข้อมูลทางสถิติที่เราเก็บรวบรวมมาทั้งหมด และยังไม่ได้นำมาดำเนินการใดๆ ทางสถิติกับข้อมูลชุดนี้ เรียกว่า **ข้อมูลดิบ (raw data)** อาจเป็นตัวเลขหรือไม่เป็นตัวเลขก็ได้ และถ้าเป็นตัวเลขอาจเป็นตัวเลขที่แสดงปริมาณหรือไม่ใช่ตัวเลขแทนปริมาณก็ได้

**การนำเสนอข้อมูล (Presentation of Data)** หมายถึง การนำข้อมูลดิบที่เก็บรวบรวมมา จัดการให้เป็นระบบตามวิธีทางสถิติ เพื่อให้ผู้อื่นหรือหน่วยงานที่เกี่ยวข้องได้ศึกษาข้อมูลนั้นๆ ได้อย่างสะดวก และรวดเร็วเป็นระเบียบแบบแผน

**การแจกแจงความถี่ของข้อมูล (frequency distribution of data)** เป็นวิธีการทางสถิติอย่างหนึ่ง โดยนำข้อมูลที่มีอยู่หรือที่เก็บรวบรวมได้(ข้อมูลดิบ) มาแบ่งออกเป็นพวกๆ(เป็นช่วงๆ) แล้วนับจำนวนข้อมูลที่มีอยู่ในแต่ละพวก ซึ่งจำนวนในแต่ละพวก เรียกว่า **ความถี่ (frequency)** และตารางที่สร้างขึ้น เรียกว่า **ตารางแจกแจงความถี่ (frequency distribution table)**

ตารางแจกแจงความถี่นิยมใช้กับข้อมูลที่มีจำนวนมาก หรือมีค่าตัวเลขซ้ำกันอยู่มาก ช่วยให้สะดวกรวดเร็วในการนำไปใช้ โดยเฉพาะอย่างยิ่งใช้ในการวิเคราะห์ข้อมูล เช่น การหาค่าเฉลี่ยของข้อมูล การหาความแปรปรวนของข้อมูล เป็นต้น ตารางแจกแจงความถี่ มี 2 วิธี คือ

1. **โดยวิธีเรียงลำดับค่าของข้อมูล** เหมาะสำหรับข้อมูลที่มีจำนวนไม่มากนัก และความห่างของข้อมูลก็ไม่มาก มีข้อมูลซ้ำกันอยู่ ตัวอย่างเช่น ในการสำรวจขนาดรองเท้าของนักเรียน 30 คน ปรากฏว่านักเรียนสวมรองเท้าเบอร์ต่างๆดังนี้

38	37	38	39	42	40	41	42	37	37
39	40	37	42	37	38	42	37	42	38
42	39	41	39	40	42	41	38	42	40

$$\text{ข้อมูลสูงสุด} = 42$$

$$\text{ข้อมูลต่ำสุด} = 37$$

$$\text{พิสัย} = \text{ข้อมูลสูงสุด} - \text{ข้อมูลต่ำสุด} = 42 - 37 = 5$$

เบอร์รองเท้า	รอยขีด	จำนวนนักเรียน
37	<del>    </del> /	6
38	<del>    </del>	5
39	////	4
40	////	4
41	///	3
42	<del>    </del> ///	8
รวม		30

2. ในกรณีที่ข้อมูลดิบเป็นตัวเลขที่แสดงปริมาณ มีจำนวนข้อมูลมากๆ และไม่ค่อยซ้ำกัน ทำโดยวิธีเรียงข้อมูลเป็นช่วงๆหรือเป็นอันตรภาคชั้น ตัวอย่างเช่น นักเรียนห้องหนึ่งสอบวิชาคณิตศาสตร์ได้คะแนนดังนี้

68 84 75 82 68 91 61 89 75 93  
 73 79 87 77 60 92 70 58 82 75  
 61 65 74 86 72 62 90 78 63 72  
 96 78 89 61 75 95 60 79 85 71  
 65 80 73 57 88 63 62 76 54 74

$$\text{ข้อมูลสูงสุด} = 96$$

$$\text{ข้อมูลต่ำสุด} = 54$$

$$\text{พิสัย} = \text{ข้อมูลสูงสุด} - \text{ข้อมูลต่ำสุด} = 96 - 54 = 42$$

$$\text{จำนวนอันตรภาคชั้น 5 ชั้น ดังนั้นแต่ละชั้นกว้าง } 42 \div 5 = 8.4 \longrightarrow 9$$

คะแนน	รอยขีด	ความถี่
54 - 62	<del>    </del> <del>    </del>	10
63 - 71	<del>    </del> ///	8
72 - 80	<del>    </del> <del>    </del> <del>    </del> //	17
81 - 89	<del>    </del> ////	9
90 - 98	<del>    </del> /	6
รวม		50

จากตารางแจกแจงความถี่ข้างต้น มีส่วนประกอบต่างๆ ดังนี้

**อันตรภาคชั้น (class interval)** หรือเรียกสั้นๆว่า **ชั้น** หมายถึงช่วงของข้อมูลในแต่ละพวกที่แบ่ง

**อันตรภาคชั้นต่ำสุด** หมายถึง อันตรภาคชั้นของข้อมูลที่มีค่าต่ำสุดอยู่ ได้แก่ อันตรภาคชั้น 54 - 62

**อันตรภาคชั้นสูงสุด** หมายถึง อันตรภาคชั้นของข้อมูลที่มีค่าสูงสุดอยู่ ได้แก่ อันตรภาคชั้น 90 - 98

**อันตรภาคชั้นต่ำกว่า** หมายถึง อันตรภาคชั้นของข้อมูลที่มีค่าน้อยกว่า ได้แก่ อันตรภาคชั้น 54 - 62

เป็นอันตรภาคชั้นที่ต่ำกว่า อันตรภาคชั้น 63 - 71

**อันตรภาคชั้นสูงกว่า** หมายถึง อันตรภาคชั้นของข้อมูลที่มีค่ามากกว่า ได้แก่ อันตรภาคชั้น 63 - 71

เป็นอันตรภาคชั้นที่ต่ำกว่า อันตรภาคชั้น 54 - 62

**ขอบล่าง (lower boundary)** ของอันตรภาคชั้น หมายถึง ค่ากึ่งกลางระหว่างค่าที่น้อยที่สุดของอันตรภาคชั้น

นั้นกับค่าที่มากที่สุดของอันตรภาคชั้นที่ต่ำกว่าหนึ่งชั้น สรุปเป็นสูตร คือ

$$\text{ขอบล่าง} = \frac{\text{ค่าที่น้อยที่สุดของอันตรภาคชั้นนั้น} + \text{ค่าที่มากที่สุดของอันตรภาคชั้นที่ต่ำกว่าหนึ่งชั้น}}{2}$$

$$\text{ได้แก่ ขอบล่างของอันตรภาคชั้น } 63 - 71 = \frac{63 + 62}{2} = 62.5$$

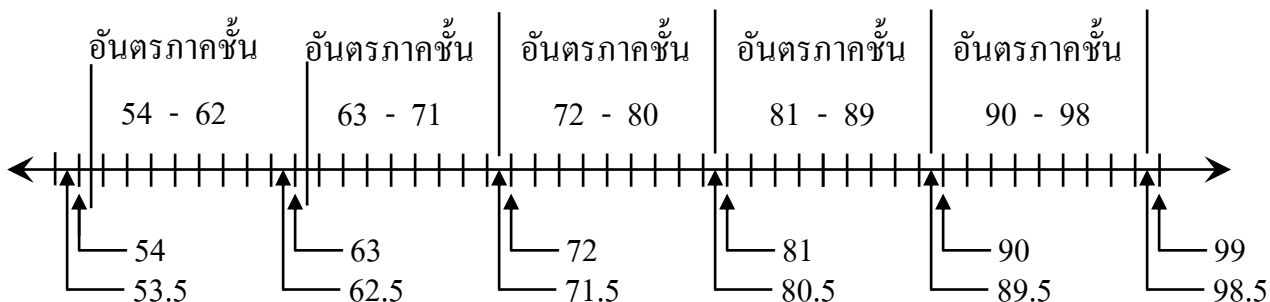
**ขอบบน (upper boundary)** ของอันตรภาคชั้น หมายถึง ค่ากึ่งกลางระหว่างค่าที่มากที่สุดของอันตรภาคชั้น

นั้นกับค่าที่น้อยที่สุดของอันตรภาคชั้นที่สูงกว่าหนึ่งชั้น สรุปเป็นสูตร คือ

$$\text{ขอบบน} = \frac{\text{ค่าที่มากที่สุดของอันตรภาคชั้นนั้น} + \text{ค่าที่น้อยที่สุดของอันตรภาคชั้นที่สูงกว่าหนึ่งชั้น}}{2}$$

$$\text{ได้แก่ ขอบล่างของอันตรภาคชั้น } 63 - 71 = \frac{71 + 72}{2} = 71.5$$

เส้นจำนวนแสดง ขอบล่าง - ขอบบน ของข้อมูล



จากรูปเส้นจำนวน จะพบว่าขอบบนของอันตรภาคชั้นหนึ่งจะเท่ากับขอบล่างของอันตรภาคชั้นที่สูงขึ้นไปหนึ่งชั้นเสมอ

**ความกว้างของอันตรภาคชั้น (Interval)** หมายถึง ผลต่างระหว่างขอบบนและขอบล่างของอันตรภาคชั้นนั้น นั่นคือ

$$\text{ความกว้างของอันตรภาคชั้น} = \text{ขอบบน} - \text{ขอบล่าง}$$

ตัวอย่างเช่น ความกว้างของอันตรภาคชั้น  $63 - 71 = 71.5 - 62.5 = 9$

**จุดกึ่งกลางชั้น (middle point) ของอันตรภาคชั้น** หมายถึง ค่ากึ่งกลางระหว่างขอบล่างและขอบบนของอันตรภาคชั้นนั้น หรือค่ากึ่งกลางระหว่างค่าที่น้อยที่สุดและค่าที่มากที่สุดของอันตรภาคชั้นนั้น นั่นคือ

$$\text{จุดกึ่งกลางชั้น} = \frac{\text{ขอบล่าง} + \text{ขอบบน}}{2}$$

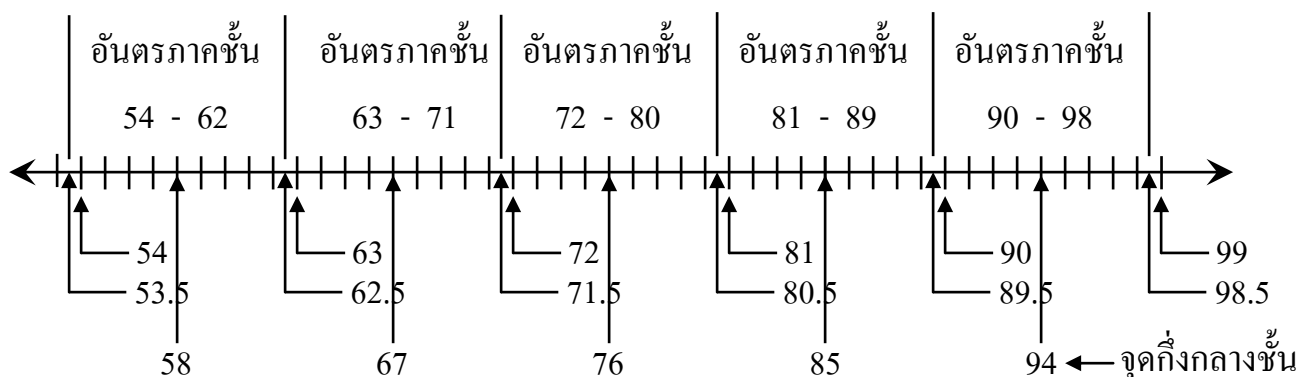
หรือ

$$\text{จุดกึ่งกลางชั้น} = \frac{\text{ค่าที่น้อยที่สุดของอันตรภาคชั้น} + \text{ค่าที่มากที่สุดของอันตรภาคชั้น}}{2}$$

ตัวอย่างเช่น จุดกึ่งกลางชั้นของอันตรภาคชั้น  $54 - 62 = \frac{53.5 + 62.5}{2} = 58$

หรือ จุดกึ่งกลางชั้นของอันตรภาคชั้น  $54 - 62 = \frac{54 + 62}{2} = 58$

**เส้นจำนวนแสดง จุดกึ่งกลางชั้น ของแต่ละอันตรภาคชั้น**



**การสร้างตารางแจกแจงความถี่**

1. ในการสร้างตารางแจกแจงความถี่ จำนวนอันตรภาคชั้นที่นิยมใช้กัน คือ 5 - 15 อันตรภาคชั้น ตามความมากน้อยของข้อมูล
2. ในตารางแจกแจงความถี่ ความกว้างของอันตรภาคชั้นไม่จำเป็นต้องเท่ากันทุกชั้นและอันตรภาคชั้นต่ำสุดหรืออันตรภาคชั้นสูงสุด อาจเป็นอันตรภาคชั้นเปิดก็ได้ เช่น เราอาจแบ่งคะแนนของนักเรียนออกเป็น 5 อันตรภาคชั้น ดังนี้ น้อยกว่าหรือเท่ากับ 50 , 51 - 59 , 60 - 68 , 69 - 77 , มากกว่าหรือเท่ากับ 78
3. ในกรณีที่มีคะแนนดิบเป็นจำนวนมากๆ ถ้าค่าที่น้อยที่สุดและค่าที่มากที่สุดของอันตรภาคชั้นเป็นค่าที่สังเกตได้ง่าย การบันทึกรอยขีดจะสะดวกขึ้น เช่น

อันตรภาคชั้น 30 - 39 , 40 - 49 , 50 - 59 , 60 - 69 ย่อมสังเกตง่ายกว่าอันตรภาคชั้น 29 - 38 , 39 - 48 , 49 - 58 , 59 - 68 เป็นต้น

4. ในกรณีที่หาความกว้างของอันตรภาคชั้นได้เป็น 9 หรือ 6 อาจเปลี่ยนเป็น 10 หรือ 5 ก็ได้ โดยที่จำนวนอันตรภาคชั้นเดิมที่กำหนดไว้ อาจเปลี่ยนไปเล็กน้อย

5. การสร้างตารางแจกแจงความถี่ นิยมสร้างให้ความกว้างของทุกอันตรภาคชั้นเท่ากัน โดยมีขั้นตอนดังนี้

5.1 หาพิสัย (range) ของข้อมูล ดังนี้

$$\text{พิสัย} = \text{ข้อมูลสูงสุด} - \text{ข้อมูลต่ำสุด}$$

5.2 หาความกว้างของอันตรภาคชั้น ดังนี้

$$\text{ความกว้างของอันตรภาคชั้น} = \text{พิสัย} \div \text{จำนวนอันตรภาคชั้น}$$

ถ้าหารลงตัวหรือหารไม่ลงตัวให้ปัดขึ้นอย่างเดียวนั้น เช่น  $20 \div 5 = 4$  ให้ปัดเป็น 5 ( $22 \div 5 = 4.4$  ปัดเป็น 5)

5.3 เขียนอันตรภาคชั้นโดยเรียงค่าน้อยไปหาค่ามาก หรือเรียงจากค่ามากไปหาค่าน้อยก็ได้ และคะแนนที่มีค่าน้อยที่สุดต้องอยู่ในอันตรภาคชั้นต่ำสุด คะแนนที่มีค่ามากที่สุดต้องอยู่ในอันตรภาคชั้นสูงสุด

5.4 หาความถี่ในแต่ละอันตรภาคชั้นจากรอยขีด (tally) ที่สำรวจได้

### ฮิสโทแกรมและรูปหลายเหลี่ยมของความถี่

การใช้กราฟแสดงการแจกแจงความถี่ทำให้เห็นการกระจายของข้อมูลได้ชัดเจนกว่าการดูจากตารางแจกแจงความถี่ แบ่งออกเป็น 2 แบบ คือ

1. **ฮิสโทแกรม (histogram)** มีลักษณะคล้ายแผนภูมิแท่งโดยจะประกอบด้วย รูปสี่เหลี่ยมมุมฉากที่วางเรียงติดต่อกันบนแกนนอน โดยมีความยาวของด้านกว้าง เท่ากับความกว้างของอันตรภาคชั้นซึ่งเท่ากันทุกชั้นและความยาวรูปสี่เหลี่ยมมุมฉากของฮิสโทแกรมจะเท่ากับความถี่ จุดปลายของด้านกว้างของรูปสี่เหลี่ยมมุมฉากแต่ละรูป คือ ขอบล่างและขอบบน

#### วิธีเขียนฮิสโทแกรม ทำดังนี้

ให้แกนนอน แทนขอบล่างและขอบบนของอันตรภาคชั้น

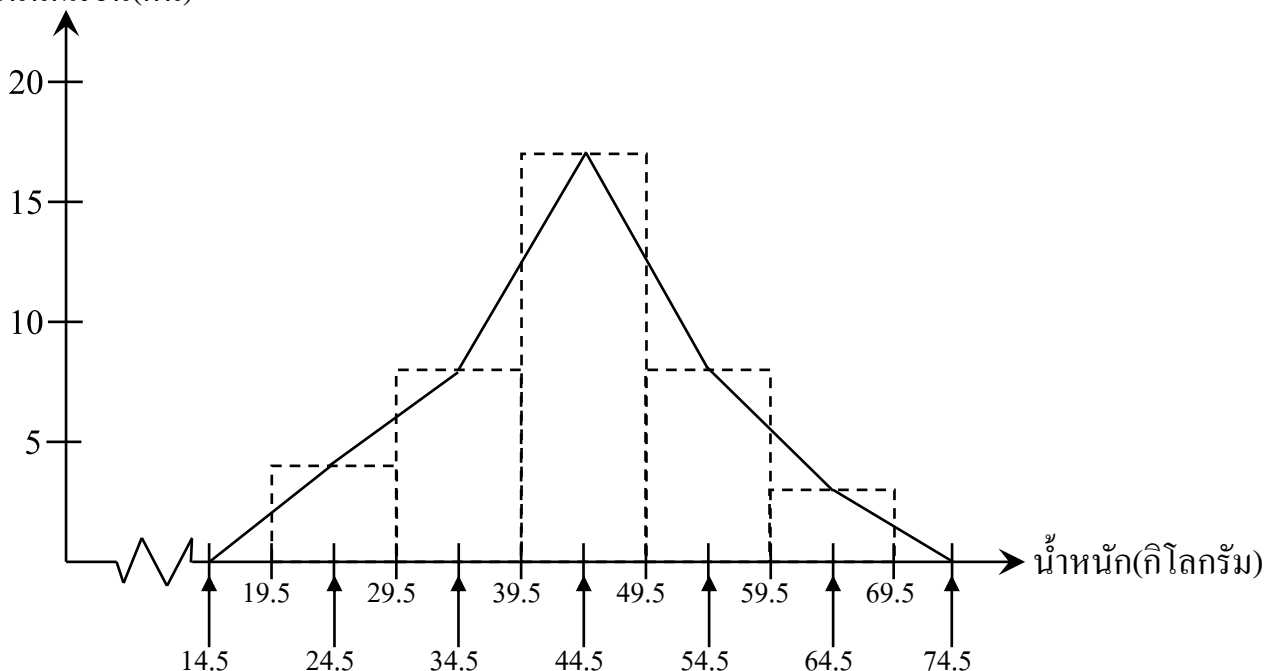
ให้แกนตั้ง แทนความถี่ของข้อมูล

ความกว้างของแท่งสี่เหลี่ยมมุมฉาก เท่ากับความกว้างของอันตรภาคชั้นซึ่งเท่ากันทุกชั้น

ความยาวของแท่งสี่เหลี่ยมมุมฉาก เท่ากับความถี่ของข้อมูลในแต่ละอันตรภาคชั้น

2. รูปหลายเหลี่ยมของความถี่ (frequency polygon) คือ รูปหลายเหลี่ยมที่ล้อมรอบด้วยแกนนอน และส่วนของเส้นตรงที่ลากเชื่อมต่อกัน ของจุดกึ่งกลางของด้านบนของรูปสี่เหลี่ยมมุมฉากแต่ละรูปของฮิสโทแกรมกับจุดกึ่งกลางชั้นของอันตรภาคชั้นที่อยู่ก่อนอันตรภาคชั้นต่ำสุดและจุดกึ่งกลางชั้นของอันตรภาคชั้นที่อยู่ถัดจากอันตรภาคชั้นสูงสุด

จำนวนนักเรียน(คน)



ข้อสังเกต พื้นที่ของรูปหลายเหลี่ยมของความถี่กับพื้นที่ของรูปสี่เหลี่ยมมุมฉากทั้งหมดในฮิสโทแกรมมีค่าเท่ากัน

### ค่ากลางของข้อมูล

ค่ากลางของข้อมูลชุดใดชุดหนึ่ง คือ ค่าที่เป็นตัวแทนของข้อมูลชุดนั้น เรานำค่ากลางของข้อมูลไปใช้ช่วยในการสรุปเรื่องราวเกี่ยวกับข้อมูลชุดนั้น ซึ่งจะช่วยให้การวิเคราะห์ข้อมูลถูกต้องมากขึ้น ค่ากลางของข้อมูล มีหลายชนิด แต่การนำไปใช้ให้ดูความเหมาะสมของข้อมูลนั้น ค่ากลางของข้อมูลที่สำคัญและนิยมใช้มี 3 ชนิด

1. ค่าเฉลี่ยเลขคณิต (arithmetic mean) ของข้อมูล คือ ค่าที่ได้จากการหารผลบวกของข้อมูลทั้งหมดด้วยจำนวนข้อมูล เมื่อข้อมูลไม่ได้แจกแจงความถี่

ถ้า  $x_1, x_2, x_3, \dots, x_N$  เป็นข้อมูล  $N$  จำนวน จะได้ว่า

$$\text{ค่าเฉลี่ยเลขคณิต} = \frac{x_1 + x_2 + x_3 + \dots + x_N}{N}$$

โดยเขียนแทนค่าเฉลี่ยเลขคณิตด้วยสัญลักษณ์  $\bar{x}$  (อ่านว่า เอ็กซ์บาร์)

$$\sum_{i=1}^N X_i = x_1 + x_2 + x_3 + \dots + x_N \text{ (อ่านว่า ซิกม่าเอ็กซ์ แทนผลบวกของข้อมูลทั้งหมด)}$$

N แทนจำนวนข้อมูลทั้งหมด

ดังนั้น

$$\bar{x} = \frac{\sum_{i=1}^N X_i}{N}$$

ข้อสังเกต 1.  $\sum_{i=1}^N X_i = N \bar{x}$

นั่นคือ ผลบวกของข้อมูลทั้งหมด = จำนวนข้อมูล  $\times$  ค่าเฉลี่ยเลขคณิต

2. ข้อมูลที่ไม่ได้แจกแจงความถี่ โดยทั่วไปจะเป็นข้อมูลที่มีจำนวนไม่เกิน 20 ตัวและเป็นข้อมูลคิ

2. **ฐานนิยม (mode)** ของข้อมูลชุดหนึ่ง คือ ข้อมูลที่มีความถี่สูงสุดในข้อมูลชุดนั้น เมื่อข้อมูลไม่ได้แจกแจงความถี่

หมายเหตุ 1. ในกรณีที่ข้อมูลทุกตัวมีความถี่เท่ากันหมดจะได้ว่าข้อมูลชุดนั้น ไม่มีฐานนิยม

2. ในกรณีที่ข้อมูลที่มีความถี่สูงสุดมีมากกว่า 1 ตัว จะได้ว่าข้อมูลทุกตัวที่มีความถี่สูงสุดนั้นเป็นฐานนิยม

3. **มัธยฐาน (median)** ของข้อมูลชุดหนึ่ง คือ ค่าที่อยู่กึ่งกลางของข้อมูลทั้งหมด ซึ่งเมื่อเรียงข้อมูลชุดนั้นจากน้อยไปมาก หรือจากมากไปน้อยแล้ว ข้อมูลที่น้อยกว่าค่านั้นจะมีจำนวนเท่ากับข้อมูลที่มากกว่าค่านั้น เมื่อข้อมูลไม่ได้แจกแจงความถี่ให้เรียงลำดับข้อมูล แล้วหาข้อมูลที่มีตำแหน่งอยู่กึ่งกลางของข้อมูลทั้งหมด ซึ่งอาจพิจารณาได้ดังนี้

ถ้าจำนวนข้อมูลเป็นจำนวนคี่ มัธยฐานจะอยู่ในตำแหน่งที่  $\frac{N+1}{2}$  เมื่อ N เป็นจำนวนข้อมูล

ถ้าจำนวนข้อมูลเป็นจำนวนคู่ มัธยฐานจะเป็นค่าเฉลี่ยของข้อมูล 2 ค่าที่อยู่ตรงกลาง คือข้อมูลในตำแหน่งที่  $\frac{N}{2}$  และ  $\frac{N}{2} + 1$  เมื่อ N เป็นจำนวนข้อมูล

การหาค่ากลางจากตารางแจกแจงความถี่

**ค่าเฉลี่ยเลขคณิต (arithmetic mean)** ของข้อมูล เมื่อข้อมูลแจกแจงความถี่แล้ว การหาค่าเฉลี่ยเลขคณิตของข้อมูล คำนวณได้จากสูตร

$$\bar{x} = \frac{\sum_{i=1}^k f_i x_i}{N}$$

เมื่อ  $f_i$  เป็นความถี่ของอันตรภาคชั้นที่  $i$

$x_i$  เป็นจุดกึ่งกลางชั้นของอันตรภาคชั้นที่  $i$

$k$  เป็นจำนวนอันตรภาคชั้น

$N$  เป็นจำนวนข้อมูลทั้งหมดหรือผลรวมของความถี่ของทุกๆอันตรภาคชั้น

**ข้อสังเกต** ในกรณีที่ความกว้างของอันตรภาคชั้นเป็น 1 จุดกึ่งกลางชั้นก็คือค่าของข้อมูลนั้นนั่นเอง

**ฐานนิยม (mode) ของข้อมูล** เมื่อข้อมูลแจกแจงความถี่แล้ว ถ้าความกว้างของอันตรภาคชั้นทุกชั้นเท่ากับ 1 ฐานนิยม คือ ข้อมูลที่มีความถี่สูงสุดนั่นเอง สำหรับกรณีอื่นๆในการเรียนระดับนี้นักเรียนจะทราบแต่เพียงว่าฐานนิยมอยู่ในอันตรภาคชั้นใดเท่านั้น

**มัธยฐาน (median) ของข้อมูล** เมื่อข้อมูลแจกแจงความถี่แล้ว ให้หาความถี่สะสมของแต่ละอันตรภาคชั้น โดยอาศัยความถี่สะสมจะทราบว่ามัธยฐานตกอยู่ในอันตรภาคชั้นใด ซึ่งถ้าความกว้างของอันตรภาคชั้นทุกชั้นเท่ากับ 1 เราจะสามารถระบุค่ามัธยฐานได้ สำหรับกรณีอื่นๆในการเรียนระดับนี้ระบุได้ว่าค่ามัธยฐานอยู่ในอันตรภาคชั้นใดเท่านั้น โดยดูจากความถี่สะสม

**ความถี่สะสมของอันตรภาคชั้นใด** คือ ผลรวมของความถี่ของอันตรภาคชั้นนั้นกับความถี่ของอันตรภาคชั้นต่ำกว่าทั้งหมด

- ข้อสังเกต**
1. ค่าเฉลี่ยเลขคณิตเป็นค่ากลางที่ได้จากการนำทุกๆค่าของข้อมูลมาเฉลี่ย แต่ฐานนิยมและมัธยฐานเป็นค่ากลางที่ใช้ข้อมูลบางค่าเท่านั้น
  2. ถ้าในข้อมูลชุดหนึ่งมีข้อมูลบางค่าที่มีค่าสูงหรือต่ำกว่าข้อมูลอื่นๆมาก จะทำให้มีผลกระทบต่อค่ากลางโดยใช้ค่าเฉลี่ยเลขคณิต กล่าวคือ อาจจะได้ค่ากลางที่มีค่าสูงหรือต่ำผิดปกติ แต่จะไม่มีผลกระทบต่อฐานนิยมหรือมัธยฐานเลย
  3. มัธยฐาน ใช้สำหรับหาค่าเฉลี่ยของรายได้ของประชากร เพราะประชากรในเมืองๆหนึ่งย่อมมีทั้งบุคคลที่มีรายได้ต่ำมากและสูงมาก ซึ่งการหาค่าเฉลี่ยโดยใช้ค่าเฉลี่ยเลขคณิตจะไม่เหมาะสม
  4. ในกรณีที่ข้อมูลเป็นข้อมูลเชิงคุณภาพ กล่าวคือ เป็นข้อมูลที่แสดงถึงสมบัติ สภาพ ฐานะ หรือความคิดเห็น เป็นต้น ข้อมูลประเภทนี้จะสามารถหาค่ากลางได้เฉพาะฐานนิยมเท่านั้น เช่น ถามว่าคนส่วนใหญ่ในตำบลหนึ่งมีอาชีพอะไร เป็นต้น



## การกระจายของข้อมูล

### ส่วนเบี่ยงเบนมาตรฐาน

#### 1. ข้อมูลที่ไม่ได้แจกแจงความถี่ (ungrouped data) เช่น

คะแนนสอบวิชาคณิตศาสตร์ของนักเรียนกลุ่มหนึ่งจำนวน 10 คน มีดังนี้

28 29 30 32 34 36 37 38 38 38

### ส่วนเบี่ยงเบนมาตรฐาน หาได้ดังนี้

#### 1. หาค่าเฉลี่ยเลขคณิต (หาผลบวกของข้อมูลทั้งหมด หารด้วยจำนวนข้อมูล) คือ

$$\frac{28 + 29 + 30 + 32 + 34 + 36 + 37 + 38 + 38 + 38}{10} = \frac{340}{10} = 34$$

#### 2. หาส่วนเบี่ยงเบนหรือหาผลต่างระหว่างแต่ละคะแนนกับค่าเฉลี่ยเลขคณิต โดยนำแต่ละคะแนนลบด้วยค่าเฉลี่ยเลขคณิต เช่น

คะแนน	ส่วนเบี่ยงเบน
28	$28 - 34 = -6$
29	$29 - 34 = -5$
30	$30 - 34 = -4$
32	$32 - 34 = -2$
34	$34 - 34 = 0$
36	$36 - 34 = 2$
37	$37 - 34 = 3$
38	$38 - 34 = 4$
38	$38 - 34 = 4$
38	$38 - 34 = 4$

3. หาคำล้งสองของส่วนเบี่ยงเบนแต่ละค่าที่หาได้จากข้อ 2

การคำนวณได้แสดงให้เห็นดังตารางต่อไปนี้

คะแนน	ส่วนเบี่ยงเบน	กำลังสองของส่วนเบี่ยงเบน
28	-6	$(-6)^2 = 36$
29	-5	$(-5)^2 = 25$
30	-4	$(-4)^2 = 16$
32	-2	$(-2)^2 = 4$
34	0	$0^2 = 0$
36	2	$2^2 = 4$
37	3	$3^2 = 9$
38	4	$4^2 = 16$
38	4	$4^2 = 16$
38	4	$4^2 = 16$

4. หาค่าเฉลี่ยเลขคณิตของกำลังสองของส่วนเบี่ยงเบน ที่ได้ในข้อ 3

ในที่นี้ค่าเฉลี่ยเลขคณิตของกำลังสองของส่วนเบี่ยงเบน เท่ากับ

$$\frac{36 + 25 + 16 + 4 + 0 + 4 + 9 + 16 + 16 + 16}{10} = \frac{142}{10} = 14.2$$

5. หารากที่สองที่เป็นบวกของค่าเฉลี่ยเลขคณิตที่ได้ในข้อ 4

ในที่นี้ได้  $\sqrt{14.2} \approx 3.8$

ดังนั้น ส่วนเบี่ยงเบนมาตรฐานของคะแนนกลุ่มนี้ มีค่าประมาณ 3.8 คะแนน

$$SD = \sqrt{\frac{\sum_{i=1}^k (x_i - \bar{x})^2}{N}}$$

ถ้า  $x_i$  แทนข้อมูลแต่ละค่า จำนวน  $N$  ตัว

## 2. ข้อมูลที่ได้มีการแจกแจงความถี่ (grouped data) เช่น

คะแนนสอบ	ความถี่ ( $f_i$ )	จุดกึ่งกลางชั้น ( $x_i$ )	$f_i x_i$
31 - 40	6	35.5	213
41 - 50	14	45.5	637
51 - 60	15	55.5	832.5
61 - 70	16	65.5	1048
71 - 80	11	75.5	830.5
81 - 90	13	85.5	1111.5
91 - 100	5	95.5	477.5
	$N = 80$		$\sum f_i x_i = 5150$

ส่วนเบี่ยงเบนมาตรฐาน หาได้ดังนี้

- หาค่าเฉลี่ยเลขคณิต (หาผลบวกของผลคูณระหว่างความถี่กับจุดกึ่งกลางชั้นทั้งหมดหารด้วยจำนวนข้อมูล) คือ

$$\bar{x} = \frac{\sum_{i=1}^k f_i x_i}{N} = \frac{5150}{80} = 64.38$$

- หาส่วนเบี่ยงเบนหรือหาผลต่างระหว่างจุดกึ่งกลางชั้นของแต่ละชั้นกับค่าเฉลี่ยเลขคณิต โดยนำจุดกึ่งกลางชั้นของแต่ละชั้นลบด้วยค่าเฉลี่ยเลขคณิต เช่น

คะแนนสอบ	ความถี่ ( $f_i$ )	จุดกึ่งกลางชั้น ( $x_i$ )	$x_i - \bar{x}$
31 - 40	6	35.5	$35.5 - 64.38 = -28.88$
41 - 50	14	45.5	$45.5 - 64.38 = -18.88$
51 - 60	15	55.5	$55.5 - 64.38 = -8.88$
61 - 70	16	65.5	$65.5 - 64.38 = 1.12$
71 - 80	11	75.5	$75.5 - 64.38 = 11.12$
81 - 90	13	85.5	$85.5 - 64.38 = 21.12$
91 - 100	5	95.5	$95.5 - 64.38 = 31.12$
	$N = 80$		

## 3. หาคำล้งสองของส่วนเบี่ยงเบนแต่ละค่าที่หาได้จากข้อ 2

การคำนวณได้แสดงให้เห็นดังตารางต่อไปนี้

คะแนนสอบ	ความถี่ ( $f_i$ )	$x_i - \bar{x}$	กำลังสองของส่วนเบี่ยงเบน
31 - 40	6	$35.5 - 64.38 = -28.88$	$(-28.88)^2 = 834.0544$
41 - 50	14	$45.5 - 64.38 = -18.88$	$(-18.88)^2 = 356.4544$
51 - 60	15	$55.5 - 64.38 = -8.88$	$(-8.88)^2 = 78.8544$
61 - 70	16	$65.5 - 64.38 = 1.12$	$(1.12)^2 = 1.2544$
71 - 80	11	$75.5 - 64.38 = 11.12$	$(11.12)^2 = 123.6544$
81 - 90	13	$85.5 - 64.38 = 21.12$	$(21.12)^2 = 446.0544$
91 - 100	5	$95.5 - 64.38 = 31.12$	$(31.12)^2 = 968.4544$
N = 80			

## 4. หาผลคูณระหว่างความถี่แต่ละชั้นกับกำลังสองของสองส่วนเบี่ยงเบนแต่ละชั้นที่หาได้จากข้อ 3

การคำนวณได้แสดงให้เห็นดังตารางต่อไปนี้

คะแนนสอบ	ความถี่ ( $f_i$ )	กำลังสองของส่วนเบี่ยงเบน	$f_i(x_i - \bar{x})^2$
31 - 40	6	$(-28.88)^2 = 834.0544$	$6 \times 834.0544 = 5004.33$
41 - 50	14	$(-18.88)^2 = 356.4544$	$14 \times 356.4544 = 4990.36$
51 - 60	15	$(-8.88)^2 = 78.8544$	$15 \times 78.8544 = 1182.82$
61 - 70	16	$(1.12)^2 = 1.2544$	$16 \times 1.2544 = 20.07$
71 - 80	11	$(11.12)^2 = 123.6544$	$11 \times 123.6544 = 1360.20$
81 - 90	13	$(21.12)^2 = 446.0544$	$13 \times 446.0544 = 5798.71$
91 - 100	5	$(31.12)^2 = 968.4544$	$5 \times 968.4544 = 4842.27$
N = 80			

5. หาค่าเฉลี่ยเลขคณิตของคุณระหว่างความถี่แต่ละชั้นกับกำลังสองของสองส่วนเบี่ยงเบนแต่ละชั้น  
ที่ได้ในข้อ 4

ในที่นี้ค่าเฉลี่ยเลขคณิตของคุณระหว่างความถี่แต่ละชั้นกับกำลังสองของสองส่วนเบี่ยงเบน  
แต่ละชั้นเท่ากับ

คะแนนสอบ	ความถี่ ( $f_i$ )	กำลังสองของส่วนเบี่ยงเบน	$f_i(x_i - \bar{X})^2$
31 - 40	6	$(-28.88)^2 = 834.0544$	$6 \times 834.0544 = 5004.33$
41 - 50	14	$(-18.88)^2 = 356.4544$	$14 \times 356.4544 = 4990.36$
51 - 60	15	$(-8.88)^2 = 78.8544$	$15 \times 78.8544 = 1182.82$
61 - 70	16	$(1.12)^2 = 1.2544$	$16 \times 1.2544 = 20.07$
71 - 80	11	$(11.12)^2 = 123.6544$	$11 \times 123.6544 = 1360.20$
81 - 90	13	$(21.12)^2 = 446.0544$	$13 \times 446.0544 = 5798.71$
91 - 100	5	$(31.12)^2 = 968.4544$	$5 \times 968.4544 = 4842.27$
	$N = 80$		$\sum_{i=1}^k f_i(x_i - \bar{X})^2 = 23198.75$

$$\bar{X} = \frac{\sum_{i=1}^k f_i(x_i - \bar{X})^2}{N} = \frac{23198.75}{80} = 289.98$$

6. หารากที่สองที่เป็นบวกของค่าเฉลี่ยเลขคณิตที่ได้ในข้อ 5

$$\text{ในที่นี้ได้ } \sqrt{289.98} \approx 17.03$$

ดังนั้น ส่วนเบี่ยงเบนมาตรฐานของคะแนนกลุ่มนี้ มีค่าประมาณ 17.03 คะแนน

$$SD = \sqrt{\frac{\sum_{i=1}^k f_i(x_i - \bar{X})^2}{N}}$$

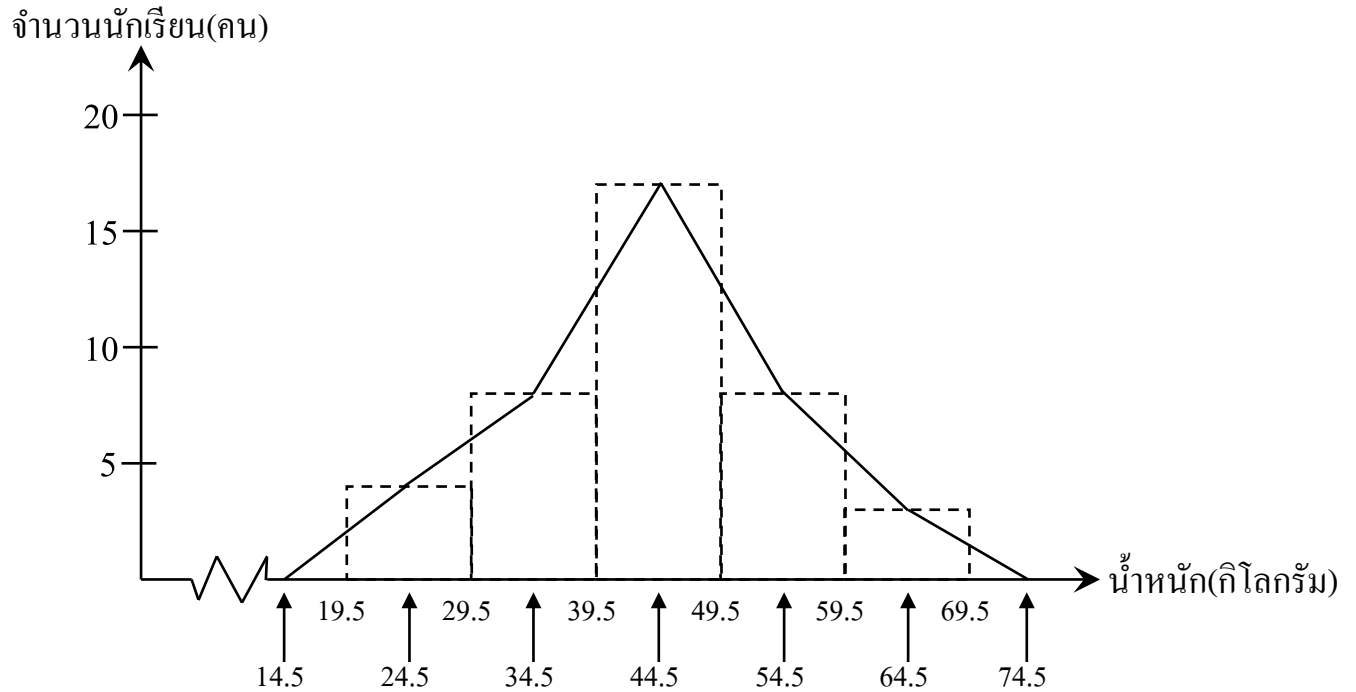
จากรายแจกแจงความถี่ของข้อมูล จำนวน  $k$  ชั้น

ให้  $x_1, x_2, x_3, \dots, x_k$  แทนจุดกึ่งกลางชั้นของข้อมูลในแต่ละชั้น

ให้  $f_1, f_2, f_3, \dots, f_k$  แทนความถี่ของข้อมูลในแต่ละชั้น

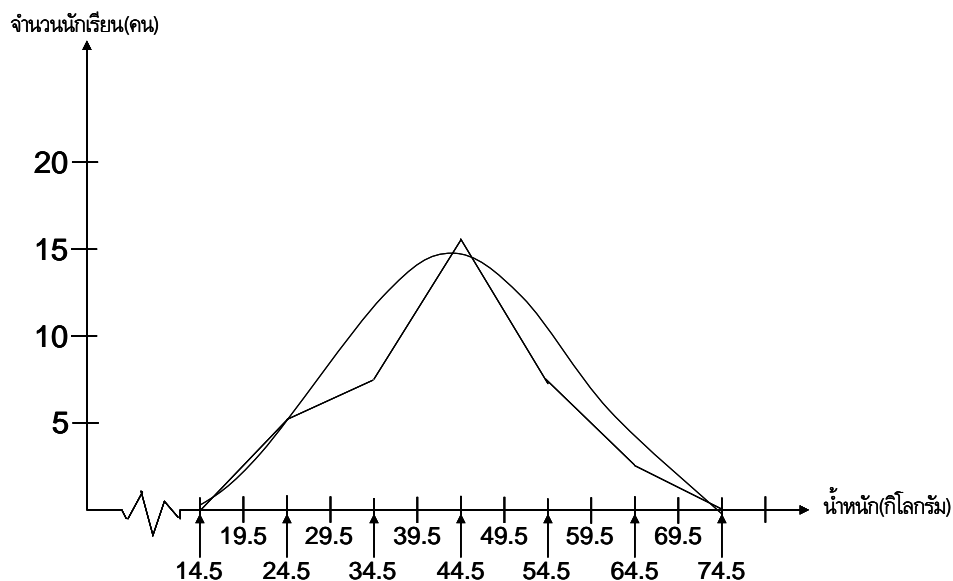
## เส้นโค้งปกติ

จากฮิสโตแกรมแสดงน้ำหนักของนักเรียน เขียนรูปหลายเหลี่ยมของความถี่ได้ดังนี้



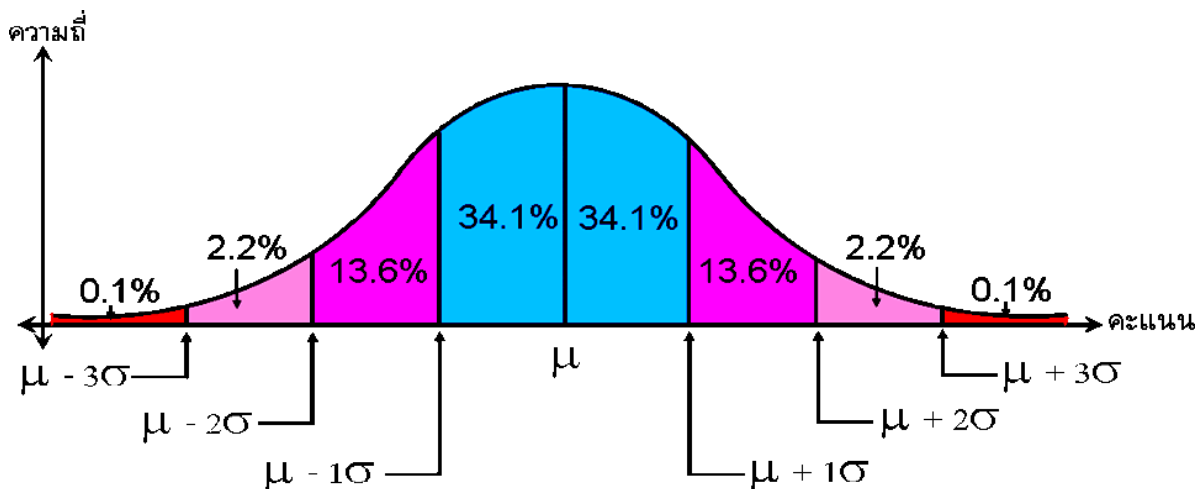
จากภาพข้างต้น พื้นที่ของรูปหลายเหลี่ยมของความถี่เท่ากับผลบวกของพื้นที่ของรูปสี่เหลี่ยมมุมฉากทุกรูปในฮิสโตแกรม

ถ้ามีจำนวนข้อมูลมากพอและกำหนดให้จำนวนของอันตรภาคชั้นมากพอ แล้วรูปหลายเหลี่ยมของความถี่ก็จะมีลักษณะใกล้เคียงกับเส้นโค้ง ดังนั้นทางสถิติจะปรับรูปหลายเหลี่ยมของความถี่ให้เป็นเส้นโค้งเรียบ โดยให้พื้นที่ภายใต้เส้นโค้งที่ปรับแล้วเท่ากับพื้นที่ของรูปหลายเหลี่ยมของความถี่รูปเดิม

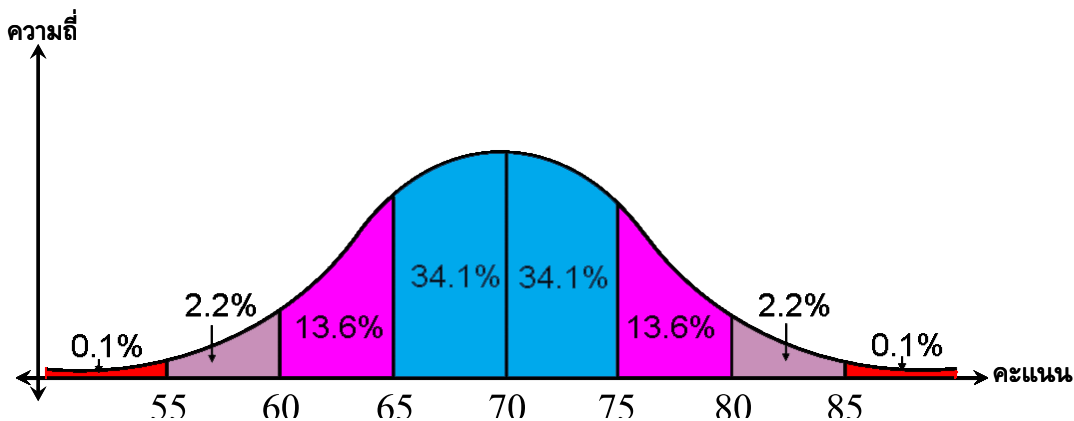


เส้นโค้งปกติมีสมบัติที่เกี่ยวข้องกับค่าเฉลี่ยเลขคณิตและส่วนเบี่ยงเบนมาตรฐาน กล่าวคือ เมื่อใช้ส่วนเบี่ยงเบนมาตรฐานเป็นหน่วยในการแบ่งแกนนอน โดยเริ่มแบ่งจากจุดที่แทนค่าเฉลี่ยเลขคณิต ไปทางขวาและทางซ้ายข้างละหนึ่งส่วนเบี่ยงเบนมาตรฐาน สองส่วนเบี่ยงเบนมาตรฐาน และสามส่วนเบี่ยงเบนมาตรฐาน ตามลำดับ ทำให้ได้ช่วงใต้เส้นโค้งปกติ 8 ช่วง แต่ละช่วงมีพื้นที่คิดเป็นเปอร์เซ็นต์ของพื้นที่ใต้เส้นโค้งทั้งหมด ดังรูป

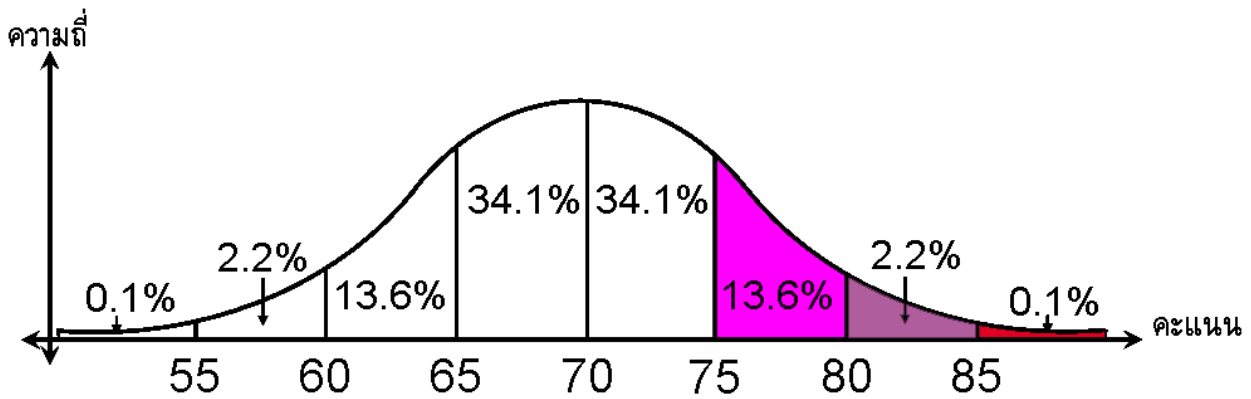
กำหนดให้                     $\mu$  (อ่านว่า มิว)                    แทนค่าเฉลี่ยเลขคณิต  
     $\sigma$  (อ่านว่า ซิกมา)                    แทนส่วนเบี่ยงเบนมาตรฐาน



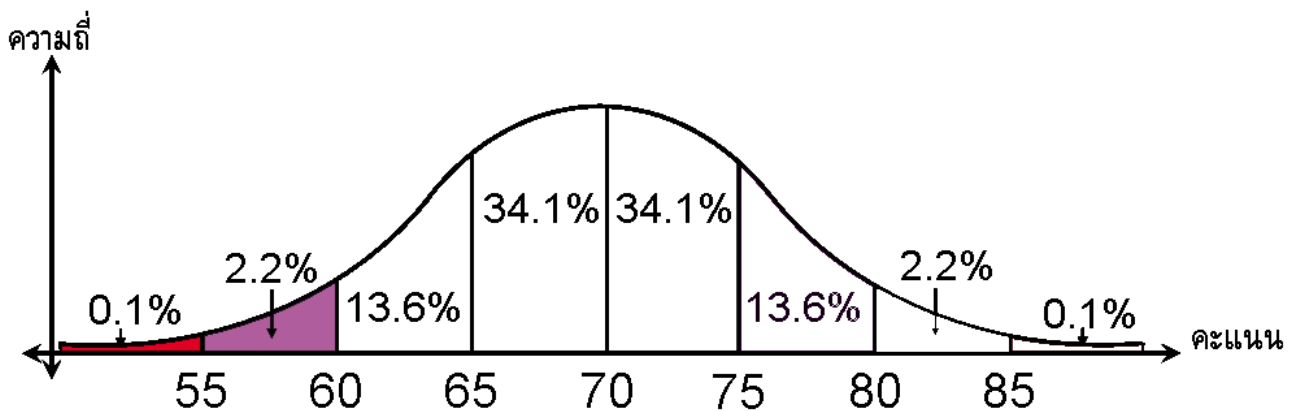
**ตัวอย่าง**                    คะแนนสอบของวิชาหนึ่งมีการแจกแจงปกติ โดยมีค่าเฉลี่ยเลขคณิตเท่ากับ 70 คะแนน และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 5 คะแนน



จากพื้นที่ใต้โค้งปกติข้างต้น เราสามารถหาพื้นที่ของช่วงคะแนนที่ต้องการได้ เช่น พื้นที่ระหว่างคะแนน 65 และ 75 เท่ากับ  $34.1 + 34.1 = 68.2 \%$



จากรูปข้างบน ผู้ที่สอบได้คะแนนมากกว่า 75 คะแนน มี  $13.6 + 2.2 + 0.1 = 15.9$  %



จากรูปข้างบน ผู้ที่สอบได้คะแนนน้อยกว่า 60 คะแนน มี  $2.2 + 0.1 = 2.3$  %

### ความคลาดเคลื่อนในการใช้สถิติ

- ความคลาดเคลื่อนในการใช้สถิติ อาจเกิดขึ้นได้ในทุกขั้นตอนของการดำเนินการทางสถิติ เริ่มตั้งแต่การเก็บรวบรวมข้อมูล การนำเสนอข้อมูล การวิเคราะห์ข้อมูล ตลอดจนถึงการแปลความหมาย
- การเก็บรวบรวมข้อมูล อาจได้ข้อมูลที่ไม่ตรงกับข้อเท็จจริง มีความลำเอียงในการเก็บรวบรวมข้อมูล หรือมีข้อผิดพลาดในการบันทึกข้อมูล
- การนำเสนอข้อมูล อาจมีการนำเสนอข้อมูลผิดความเป็นจริงหรือนำเสนอข้อมูลแล้วทำให้ผู้ใช้เกิดการเข้าใจผิด
- การวิเคราะห์ข้อมูล อาจมีการใช้สถิติที่ไม่เหมาะสมกับข้อมูล เช่น ในชุดของข้อมูลที่ใช้มาตรฐานเพื่อบอกค่ากลาง แต่กลับใช้ค่าเฉลี่ยเลขคณิต
- การแปลความหมาย อาจมีการอ่านข้อมูลที่น่าเสนอผิดความจริงหรือมีการตีความเกินความเป็นจริง