

บทที่ 6

การวิเคราะห์ข้อมูลเบื้องต้น

CR. JIRAPORN SANJUN

6.1 ข้อมูลและการจัดระเบียบข้อมูล

6.1.1 คำศัพท์เกี่ยวกับข้อมูล

6.1.2 ขั้นตอนวิธีทางสถิติ

6.1.3 การจัดระเบียบข้อมูล

6.1.4 การนำเสนอข้อมูล

6.1.1 คำศัพท์เกี่ยวกับข้อมูล

ข้อมูลทางสถิติ

ประชากร ตัวอย่าง

ประเภทข้อมูล ข้อมูลเชิงปริมาณ ข้อมูลเชิงคุณภาพ

แหล่งข้อมูล แหล่งปฐมภูมิ แหล่งทุติยภูมิ



6.1.1 คำศัพท์เกี่ยวกับข้อมูล

ข้อมูล คือ ข้อเท็จจริงที่อยู่ในรูปของจำนวน
ข่าวสาร ข้อความ หรือความคิดเห็น ใดๆ
อย่างหนึ่งหรือหลายอย่างเพื่อใช้ในการศึกษา
หรือสรุปเรื่องราวที่เราสนใจ

6.1.1 คำศัพท์เกี่ยวกับข้อมูล

ประเภทข้อมูล

1. ข้อมูลเชิงคุณภาพ แสดงลักษณะสิ่งที่สนใจ
เพศ สถานะภาพ สาขาวิชา
ความคิดเห็น เชื้อชาติ
2. ข้อมูลเชิงปริมาณ บอกขนาดหรือปริมาณ
คะแนนสอบ อายุ น้ำหนัก รายได้

กิจกรรม

1. จงพิจารณาข้อมูลต่อไปนี้ว่าเป็นข้อมูลเชิงปริมาณหรือข้อมูลเชิงคุณภาพ และเติมคำในช่องว่าง
.....1) หมายเลขโทรศัพท์
.....2) ความสูงของนักเรียน
.....3) จำนวนเงินที่นักเรียนนำมาโรงเรียน
.....4) ทะเบียนรถยนต์
.....5) อาชีพของผู้ปกครอง
.....6) เพศของนักเรียน
.....7) บ้านเลขที่
.....8) ราคาน้ำมัน
.....9) น้ำหนักของนักเรียน
.....10) วุฒิการศึกษา

6.1.1 คำศัพท์เกี่ยวกับข้อมูล

แหล่งข้อมูล

แหล่งข้อมูล หมายถึง แหล่งที่ได้มาซึ่งข้อมูลทางสถิติ แบ่งได้ 2 ประเภท

1. แหล่งปฐมภูมิ (PRIMARY SOURCE) เป็นแหล่งข้อมูลที่จัดเก็บรวบรวมข้อมูลด้วยตนเอง

ข้อดี ได้ข้อมูลโดยตรงตามที่ต้องการ

ข้อเสีย ใช้เวลาและงบประมาณค่อนข้างมาก

2. แหล่งทุติยภูมิ (SECONDARY SOURCE) เป็นแหล่งข้อมูลที่ไม่ได้จัดเก็บข้อมูลครั้งแรกเอง ข้อมูลที่รวบรวมจะได้อาจมาจากข้อมูลที่ผ่านมาผ่านการรวบรวมไว้ก่อนแล้ว

ข้อดี ซึ่งจะทำให้ประหยัดเวลาและค่าใช้จ่ายจำนวนมาก

ข้อเสีย อาจจะไม่ได้อินโฟรมาตามที่ต้องการอย่างแท้จริง

6.1.1 คำศัพท์เกี่ยวกับข้อมูล

กลุ่มข้อมูลที่เราจะศึกษาแบ่งได้เป็น 2 กลุ่มซึ่งได้แก่ ประชากร และ ตัวอย่าง

ประชากร หมายถึง ทุกหน่วยที่ให้รายละเอียดในเรื่องที่ต้องการศึกษา

ตัวอย่าง หมายถึง ส่วนหนึ่งของประชากร

ข้อสังเกตสำหรับสมาชิกของตัวอย่างคือ สมาชิกแต่ละตัวในตัวอย่างจะต้องได้มาโดยการสุ่ม โดยมีข้อตกลงว่า สมาชิกแต่ละหน่วยจะต้องมีโอกาสที่จะถูกเลือกจากประชากรเท่า ๆ กัน เพื่อให้มั่นใจว่ากลุ่มตัวอย่างนั้นเป็นตัวแทนที่ดีของประชากรนั้น

6.1.2 ขั้นตอนวิธีทางสถิติ

การเก็บรวบรวมข้อมูล (DATA COLLECTION)

การนำเสนอข้อมูล (DATA PRESENTATION)

การวิเคราะห์ข้อมูล (DATA ANALYSIS)

การตีความหมายข้อมูล (DATA INTERPRETATION)

6.1.2 ขั้นตอนวิธีทางสถิติ

การเก็บรวบรวมข้อมูล เป็นการวางแผนเพื่อศึกษาเรื่องใดเรื่องหนึ่ง แล้วจึงทำการเก็บรวบรวมข้อมูลด้วยวิธีต่าง ๆ ได้แก่ การสังเกต การสัมภาษณ์ การทำแบบสอบถาม การทำแบบทดสอบ

การนำเสนอข้อมูล เป็นการนำข้อมูลที่เก็บรวบรวมมาได้มาจัดการให้อยู่ในรูปแบบที่สามารถเข้าใจได้ง่าย

การวิเคราะห์ข้อมูล เป็นการนำข้อมูลที่เก็บรวบรวมมาได้มาคำนวณหาค่าทางสถิติ เพื่อตอบคำถามในเรื่องที่กำลังศึกษา

การตีความหมายของข้อมูล เป็นการอธิบายผลที่ได้จากการวิเคราะห์ข้อมูลเพื่อให้เกิดความเข้าใจ

ตัวอย่าง

อาสาสมัครคนหนึ่ง ต้องการศึกษาสภาวะโภชนาการของเด็กใน
เบื้องต้นจะต้องทราบน้ำหนักของเด็ก เพื่อตอบคำถามต่อไปนี้
น้ำหนักของเด็กที่น้อยที่สุดเท่ากับเท่าไร
น้ำหนักของเด็กที่มากที่สุดเท่ากับเท่าไร
ส่วนใหญ่น้ำหนักเด็กมีน้ำหนักเท่าไร

จากการสำรวจน้ำหนักของเด็ก 40 คน (หน่วย : กิโลกรัม) ดังนี้

30	32	40	25	36	35	23	27	26	34
35	32	42	36	34	25	31	28	33	24
24	35	40	35	36	42	28	26	35	34
27	34	28	41	36	32	38	35	26	24

วิธีทำ

1) การจัดเรียงข้อมูลตามลำดับ หมายถึง การนำเอาข้อมูลดิบมาจัดเรียงลำดับจากมากไปหาน้อยหรือน้อยไปหามาก

23	24	24	25	25	26	26	27	27	27
28	28	28	30	31	32	32	32	33	34
34	34	34	35	35	35	35	35	35	36
36	36	36	38	38	40	40	41	42	42

2). การแจกแจงความถี่

การแจกแจงความถี่ เป็นการเรียงลำดับจากน้อยไปหามากหรือจากมากมาหาน้อย โดยมีการระบุจำนวนที่ซ้ำกันของข้อมูลแต่ละค่าซึ่งเรียกว่าความถี่ (FREQUENCY) การเรียงลำดับอาจจะเรียงทีละค่า หรืออาจจะจัดเรียงเป็นกลุ่มก็ได้เรียกว่า การจัดเรียงเป็นกลุ่ม (GROUP DATA) นอกจากนี้รูปแบบของการแจกแจงความถี่อาจจะทำได้หลายประเภท เช่น ตารางแจกแจงความถี่ ตารางแจกแจงความถี่สะสม ตารางแจกแจงความถี่สัมพัทธ์ และตารางแจกแจงความถี่สะสมสัมพัทธ์

การแจกแจงความถี่ทำได้ 2 อย่างคือ การแจกแจงความถี่ที่จัดเรียงข้อมูลทีละค่า และการแจกแจงความถี่ด้วยการจัดเรียงเป็นกลุ่ม ซึ่งมีวิธีสร้างดังนี้

2.1 ตารางแจกแจงความถี่ที่จัดเรียงข้อมูลที่ละค่า

ตารางแบบนี้เป็นตารางที่แสดงให้เห็นว่าข้อมูลแต่ละค่ามีปรากฏซ้ำจำนวนเท่าไร มีวิธีสร้างเป็นขั้น ๆ ดังนี้

ขั้นที่ 1 หาค่าต่ำสุด และสูงสุดจากข้อมูลที่กำหนดให้ เช่น ค่าต่ำสุดเท่ากับ 29 และค่าสูงสุดเท่ากับ 45

ขั้นที่ 2 เรียงข้อมูลจากต่ำสุดไปสูงสุด หรือจากสูงสุดไปต่ำสุด ค่าที่ไม่มีปรากฏในชุดข้อมูล ก็ไม่จำเป็นต้องบันทึกไว้

ขั้นที่ 3 สร้างรอยคະແນน โดยมีขีดรอยคະແນนของแต่ละค่า จนครบตามกำหนดข้อมูลทั้งหมด แล้วรวมรอยคະແນนใส่ไว้ในช่องความถี่

น้ำหนัก	รอยขีด	ความถี่
45	/	1
42	//	2
41	/	1
40	////	4
39	/	1
37	/	1
36	/	1
35	//	2
32	/	1
31	///	3
30	/	1
29	/	1
รวม		20

2.2 ตารางแจกแจงความถี่เป็นกลุ่ม

การสร้างตารางแจกแจงความถี่แบบนี้ หมายถึงการสร้าง ตารางโดยแบ่งค่าของข้อมูลเป็นช่วงซึ่งเรียกว่า *อันตรภาคชั้น* (CLASS INTERVAL) แต่ละชั้นประกอบด้วยค่าที่เป็นไปได้ หลาย ๆ ค่า ช่วงแต่ละช่วงอาจจะเท่ากันหรือไม่เท่ากันก็ได้ แต่ใน ทางปฏิบัติจะหลีกเลี่ยงการสร้างตารางแจกแจงความถี่ที่มี อันตรภาคชั้นไม่เท่ากัน ทั้งนี้เพราะการตีความจากตารางจะทำได้ ลำบากเนื่องจากต้องหาเหตุผลสนับสนุนการแบ่งช่วงของข้อมูล ดังกล่าวด้วย

ตารางแจกแจงความถี่ที่จัดเรียงเป็นอันดับจากชั้นผลการสอบวิชาคณิตศาสตร์
ของนักศึกษาปี 1 จำนวน 60 คน ดังนี้

คะแนน	ความถี่
25- 34	2
35 – 44	6
45 – 54	13
55 – 64	28
65 – 74	11
รวม	60

ช่วง 25 - 34, 35 - 44, ..., 65 - 74 เรียกว่า
 อัตรภาคชั้น (CLASS INTERVAL)
มีจำนวนอัตรภาคชั้น ทั้งหมด 5 อัตรภาคชั้น

ในอัตรภาคชั้น 35 - 44 ประกอบด้วย 35 36 37 38
39 40 41 42 43 44 มีทั้งหมด 10 จำนวน เรียกว่ามี
 ความกว้างเท่ากับ 10

35 เป็น **ขีดจำกัดล่าง (LOWER LIMIT)** ของอัตรภาค
 ชั้น 35 - 44

44 เป็น **ขีดจำกัดบน (UPPER LIMIT)** ของอัตรภาคชั้น
 35 - 44

จุดกึ่งกลางชั้น (MID POINT) และหาได้จากสูตร

$$\text{จุดกึ่งกลางชั้น} = \frac{\text{ขีดจำกัดบน} + \text{ขีดจำกัดล่าง}}{2}$$

เช่น จุดกึ่งกลางชั้นของชั้น

$$\begin{aligned} 35 + 44 &= \frac{44 + 35}{2} \\ &= 39.5 \end{aligned}$$

ในกรณีที่เป็นการแจกแจงความถี่ข้อมูลชนิดต่อเนื่อง เช่น น้ำหนัก ส่วนสูง เวลา อื่น ๆ ค่าที่วัดได้มักเป็นค่าใกล้เคียงเช่น น้ำหนัก 40 กิโลกรัม ค่าจริงอาจจะมากกว่าหรือน้อยกว่า 40 กิโลกรัม ซึ่งประมาณได้ว่าเป็นค่าที่อยู่ระหว่าง 39.5 ถึง 40.5 ดังนั้นการแจกแจงความถี่จึงต้องกำหนดขอบเขตที่มีลักษณะต่อเนื่องครอบคลุม ข้อมูลที่จัดด้วยระบบนี้เรียกว่า ขีดจำกัดบนจริง และขีดจำกัดล่างจริง (TRUE UPPER LIMIT AND TRUE LOWER LIMIT) โดยใช้จุดที่อยู่ตรงกลางระหว่างค่าที่น้อยที่สุดในอันตรภาคชั้นนั้นกับค่าที่มากที่สุดของชั้นที่อยู่ติดกัน เป็นขีดจำกัดล่างและขีดจำกัดบนของชั้นที่อยู่ติดกันเช่น จากตัวอย่าง

จุดที่อยู่ตรงกลางของ 34 และ 35 ของชั้น 25 - 34 กับ 35 - 44 คือ 34.5

จะได้ว่า 34.5 เป็นขีดจำกัดบนจริงของชั้น 25 - 34

และ 34.5 เป็นขีดจำกัดล่างจริงของชั้น 35 - 44

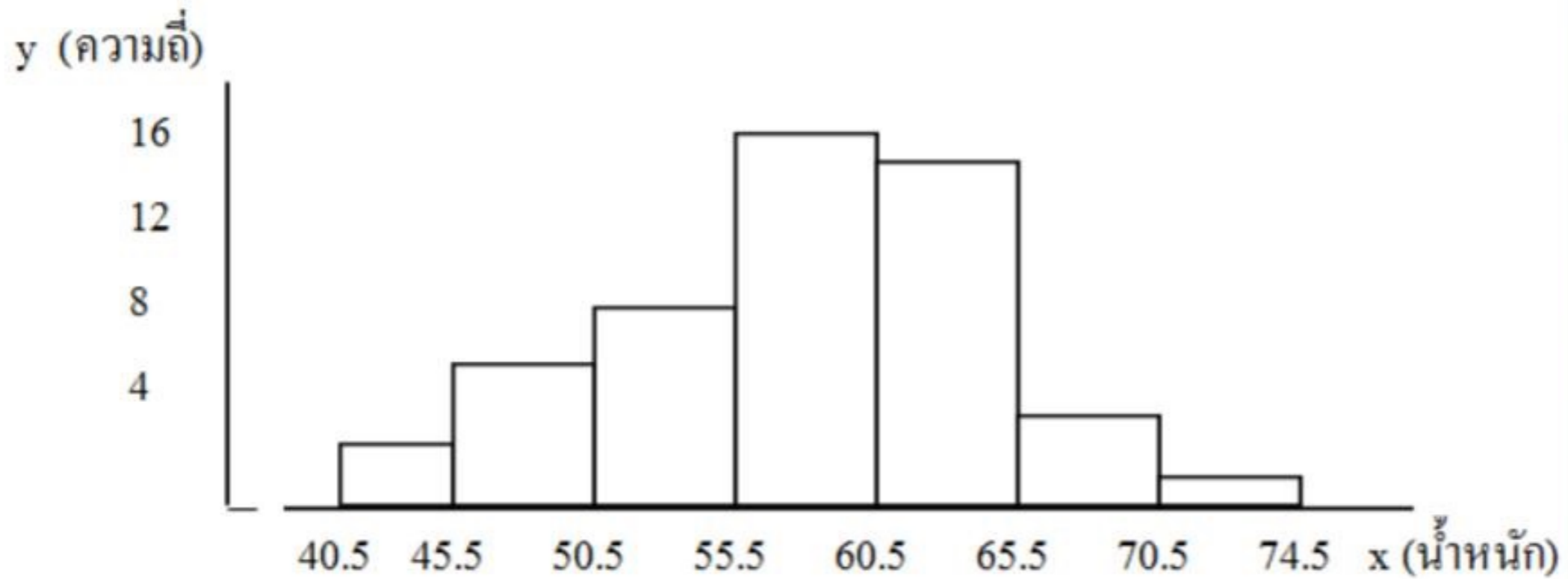
3 แผนภูมิและกราฟของการแจกแจงความถี่

แผนภูมิและกราฟของการแจกแจงความถี่ เป็นการแสดงการแจกแจงข้อมูลให้มีลักษณะเป็นรูปธรรมขึ้น ซึ่งมีวิธีต่าง ๆ ดังนี้

ฮิสโตแกรม เป็นวิธีแสดงการแจกแจงความถี่ด้วยแผนภูมิแท่ง มีวิธีสร้าง 2 วิธีคือ

วิธีที่ 1 แสดงจำนวนความถี่ของแต่ละอันตรภาคชั้นตามแนวแกน Y ด้วยความยาวของแท่งสี่เหลี่ยมผืนผ้า และความกว้างของช่วงคะแนนแทนความกว้างของแท่งสี่เหลี่ยมแสดงตามแนวแกน X

จากตัวอย่าง สร้างฮิสโตแกรม ได้ดังนี้

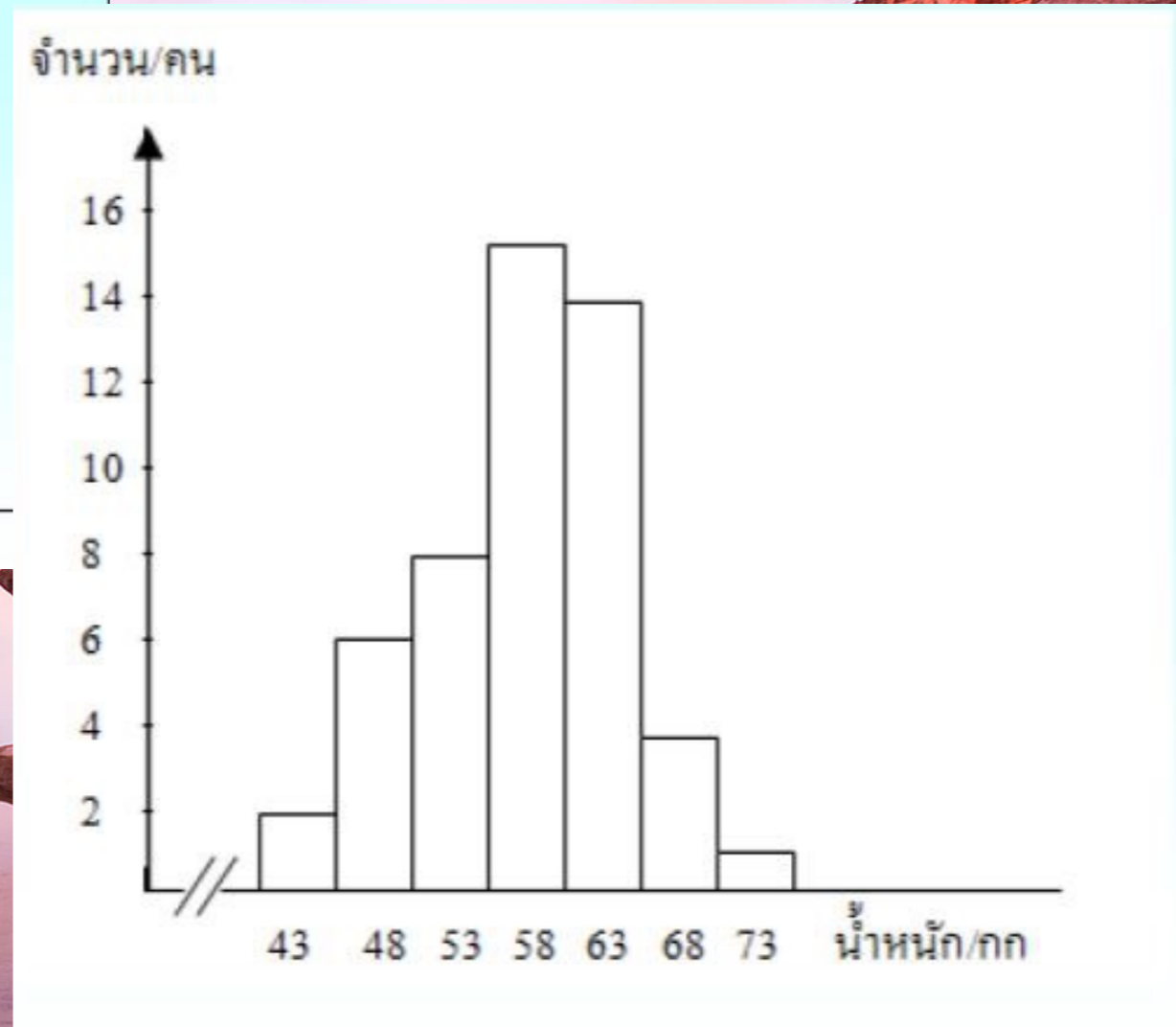


วิธีที่ 2 กำหนดจุดกลางชั้นของแต่ละอันตรภาคชั้น บนแกน X และ ความกว้างของแต่ละชั้นเทียบกับความกว้าง ของอันตรภาคชั้นในกรณีที่มีความกว้างของแต่ละอันตรภาคชั้นไม่เท่ากัน จะแสดงความถี่ของแต่ละอันตรภาคชั้นด้วยความยาวของแต่ละชั้นเหลี่ยม แต่ถ้าในกรณีความกว้างของแต่ละอันตรภาคชั้นเท่ากัน จะแสดงความถี่ของแต่ละอันตรภาคชั้นด้วยพื้นที่ของรูปสี่เหลี่ยม โดยกำหนดให้ความกว้างของแต่ละชั้นเหลี่ยมเท่ากับ 1 หน่วย ดูจากตัวอย่าง

จากตารางแจกแจงความถี่ในตัวอย่าง มีวิธีสร้างฮิสโตแกรมเป็นชั้น ๆ ดังนี้

1. หาจุดกลางชั้นของแต่ละอันตรภาคชั้น

ช่วงคะแนน	จุดกลางชั้น	ความถี่
41 – 45	43	2
46 – 50	48	6
51 – 55	53	8
56 – 60	58	15
61 – 65	63	14
66 – 70	68	4
71 – 75	73	1



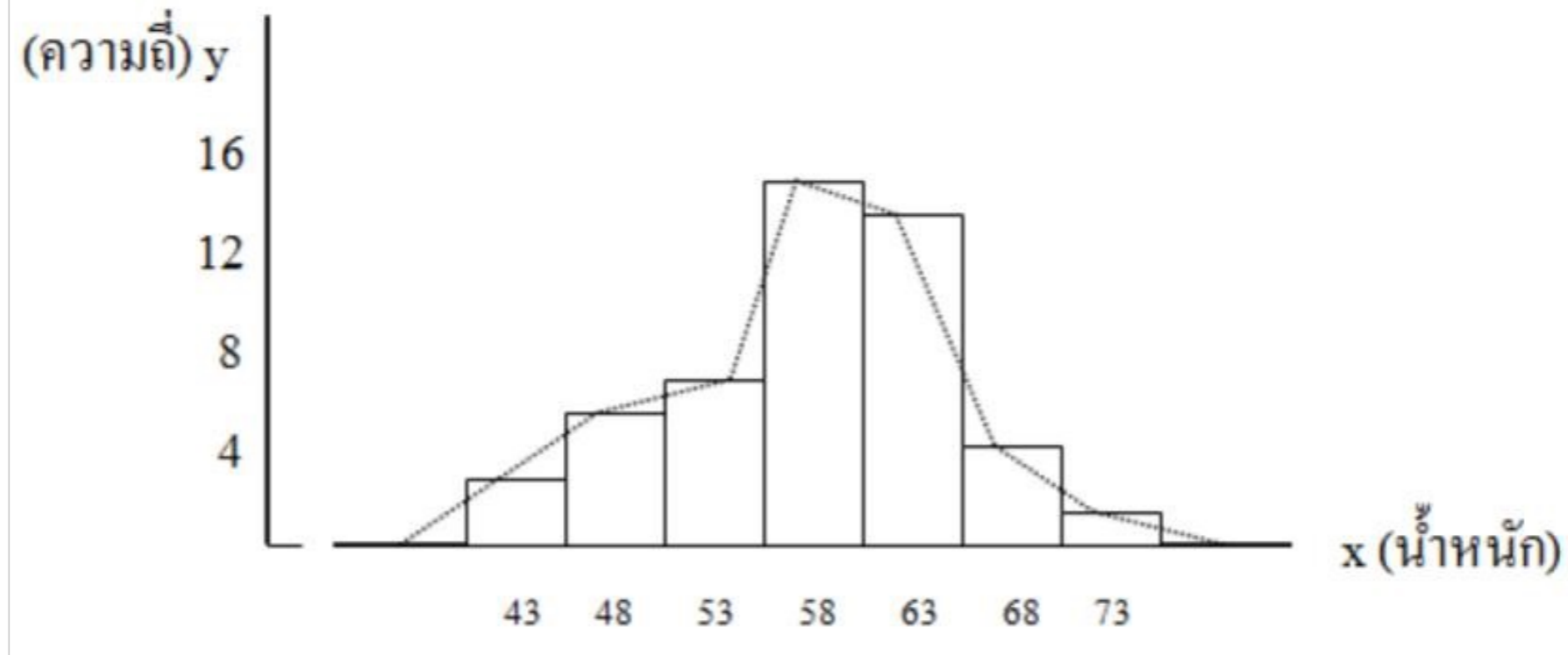
2. แทนจุดกึ่งกลางชั้นบนแกน X แล้วลากเส้นจากจุดกึ่งกลางชั้นตั้งฉากกับแกน X ยาวเท่ากับจำนวนความถี่ของแต่ละอันดับชั้น

3. ใช้แนวเส้นตรงที่ลากจากจุดกลางชั้นเป็นแนวกลางของแต่ละสี่เหลี่ยมสร้างแท่งสี่เหลี่ยมผืนผ้า ให้มีความกว้างเท่ากับความกว้างของอันดับชั้น

รูปหลายเหลี่ยมความถี่

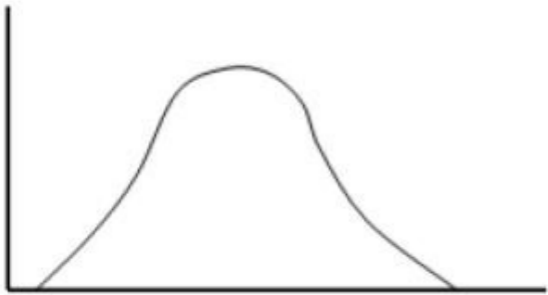
รูปหลายเหลี่ยมความถี่เป็นรูปที่เกิดขึ้นจากการลากเส้นเชื่อมต่อระหว่างจุดกึ่งกลางของปลายแท่งสี่เหลี่ยมผืนผ้าของฮิสโตแกรม โดยเริ่มตั้งแต่จุดกลางชั้นของชั้นที่ขยายทางด้านค่าต่ำสุดถึงชั้นที่ขยายออกทางด้านค่าสูงสุดอย่างละ 1 ชั้น จากตัวอย่าง 7 จุดกลางชั้นของชั้นที่ขยายออกทั้งทางด้านค่าต่ำสุด และสูงสุด คือ 38 และ 78 ตามลำดับ จะมีวิธีสร้างดังตัวอย่างต่อไปนี้

จากฮิสโตแกรมในตัวอย่าง สร้างรูปหลายเหลี่ยมความถี่ได้ ดังนี้



โค้งความถี่มีคุณสมบัติในการแสดงลักษณะโดยส่วนรวมของข้อมูลแต่ละชุด รูปร่างของโค้งความถี่จะแตกต่างกันออกไปตามลักษณะของการแจกแจงของข้อมูลแต่โดยปกติแล้วรูปแบบของการแจกแจงที่พบบ่อยมาก จะมีลักษณะต่าง ๆ ดังนี้

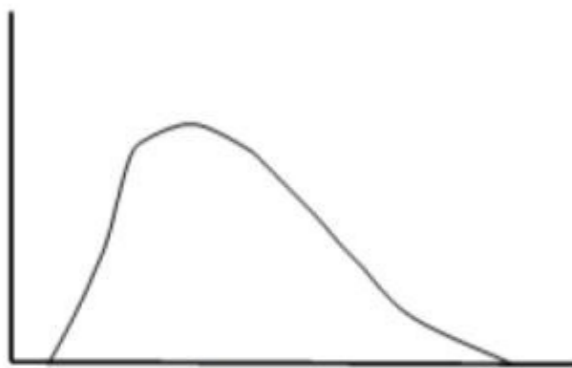
1. โค้งรูปประฆัง



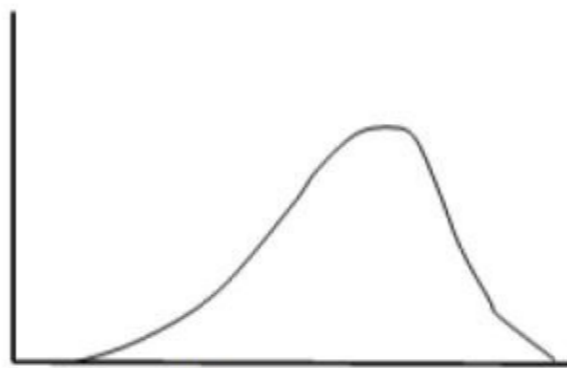
โค้งรูปประฆังว่า

ส่วนสูงที่สุดจะอยู่ในตำแหน่งตรงกลาง ค่าต่ำลงมาจะอยู่ด้านซ้ายและขวา เราจะเรียกว่า **โค้งสมมาตร (SYMMETRICAL CURVE)**

2. โค้งเบ้



เบ้ขวา



เบ้ซ้าย

โค้งเบ้ขวา

ข้อมูลส่วนใหญ่จะอยู่ทางคะแนนน้อย

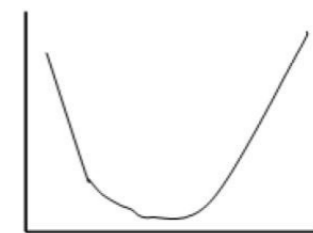
โค้งเบ้ซ้าย

ข้อมูลส่วนใหญ่จะอยู่ทางคะแนนมาก

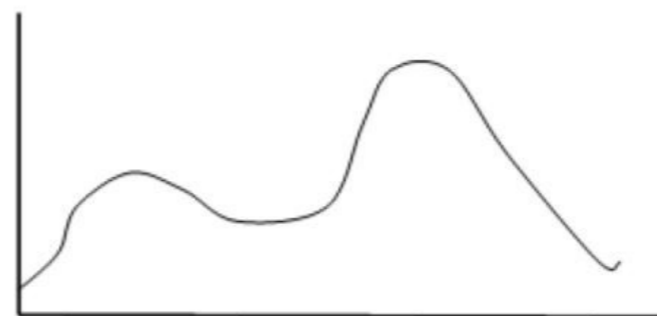
3. โค้งรูปตัว J (J - shaped curve)



4. โค้งรูปตัว U (U - shaped curve)



5. โค้งที่มีหลายยอด (multimodal curve)



การนำเสนอข้อมูล

การนำเสนอข้อมูลเป็นขั้นตอนหนึ่งของสถิติ เพื่อเป็นการเตรียมข้อมูลให้พร้อมสำหรับการวิเคราะห์และการแปลความหมาย สรุปผลข้อมูล โดยทั่วไปการนำเสนอข้อมูลสถิติอาจทำได้ใน 2 ลักษณะใหญ่ ๆ คือ

1. การนำเสนอข้อมูลอย่างไม่เป็นแบบแผน (INFORMAL PRESENTATION)
หมายถึง การนำเสนอข้อมูลที่ไม่มีการกำหนด หรือแบบแผนที่แน่นอนตายตัว ที่นิยมใช้มีสองวิธีคือ การนำเสนอข้อมูลในรูปบทความหรือข้อความเรียง และการนำเสนอข้อมูลในรูปบทความกึ่งตาราง

1.1 การนำเสนอข้อมูลในรูปบทความ (TEXT PRESENTATION)

การนำข้อมูลมาเสนอเป็นส่วนหนึ่งของข้อความ โดยทั่วไปจะพบมากในบทสรุป ที่ใช้เวลาในการรายงานไม่มากนัก

เช่น ในปีงบประมาณ 2534 ได้ผลิตบัณฑิตสาขาวิชาการศึกษา จำนวน 12,359 คน คิดเป็นร้อยละ 47.22 ผลิตบัณฑิตสาขาวิชาวิทยาศาสตร์ 5,655 คน คิดเป็นร้อยละ 21.61 และผลิตบัณฑิตสาขา วิชาศิลปศาสตร์ 8,159 คน คิดเป็นร้อยละ 31.17

1.2 การนำเสนอข้อมูลในรูปข้อความกึ่งตาราง (SEMI - TABULAR ARRANGEMENT)

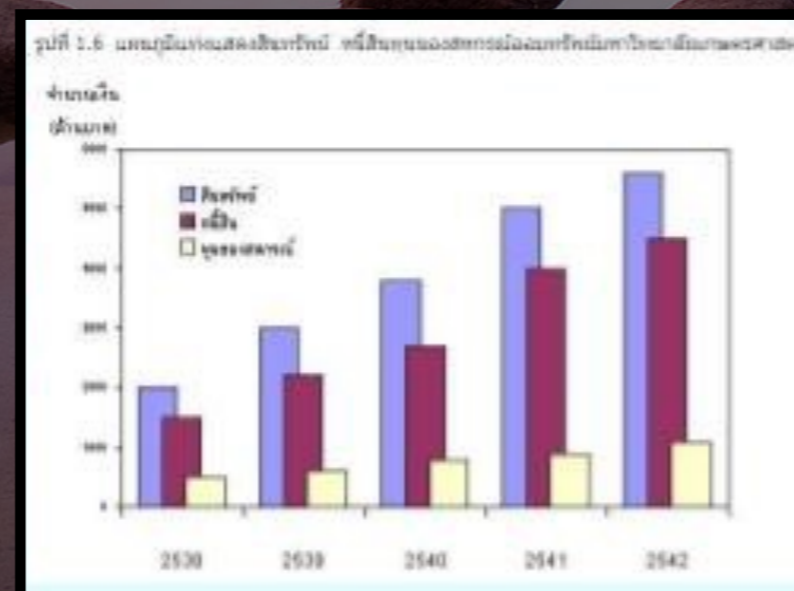
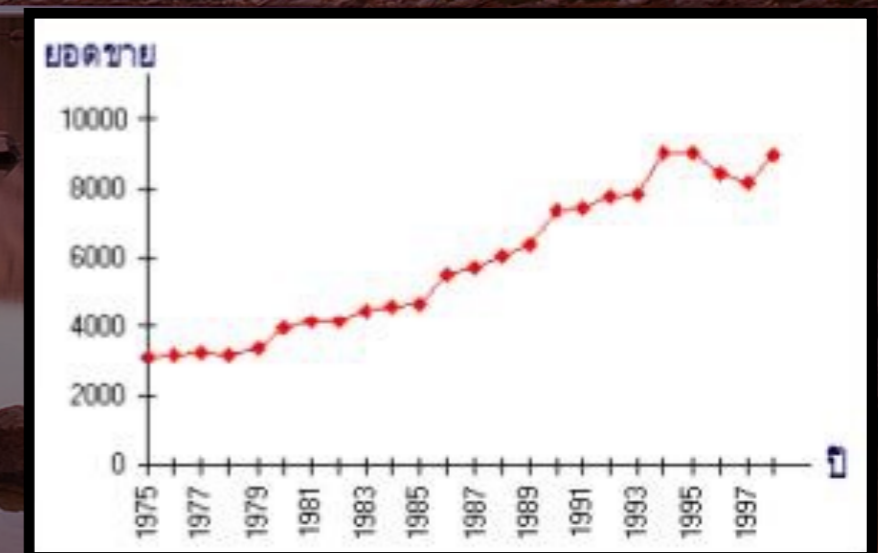
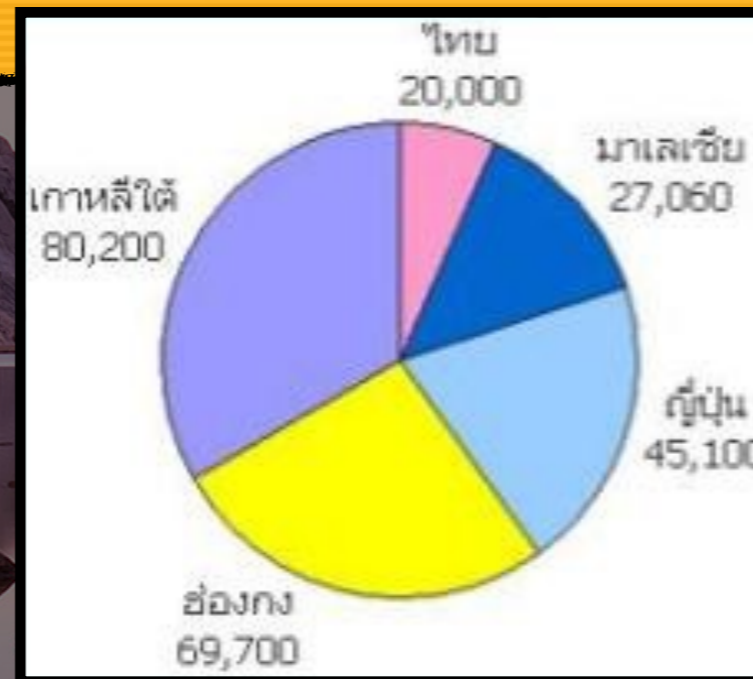
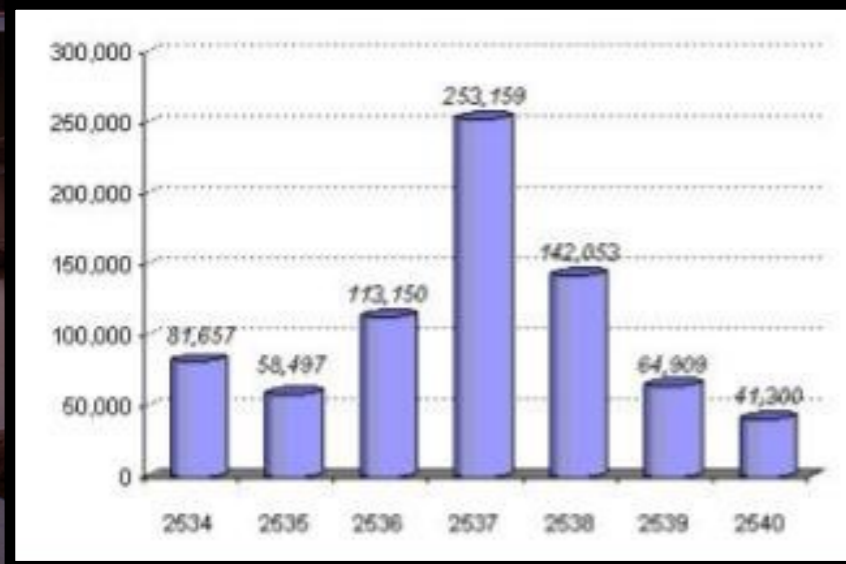
คือการนำเสนอข้อมูล โดยแยกตัวเลขออกจากข้อความ เพื่อต้องการให้เห็นตัวเลขที่ชัดเจนและ เปรียบเทียบความแตกต่างได้สะดวกยิ่งขึ้น ตัวอย่าง

บริษัทคอมพิวเตอร์แห่งหนึ่ง มีจำนวนยอดขายประจำเดือนมิถุนายน ของลูกค้าจำแนกตามภาคต่าง ๆ ดังนี้

ภาค	จำนวนยอดขาย (พันเครื่อง)
เหนือ	210
กลาง	398
ตะวันออก	135
ตะวันออกเฉียงเหนือ	102
ใต้	170

2. การนำเสนอข้อมูลอย่างเป็นแบบแผน (FORMAL PRESENTATION)

หมายถึง การนำเสนอ ข้อมูลที่มีกฎเกณฑ์ซึ่งจะต้องปฏิบัติตามมาตรฐานที่กำหนดไว้เป็นแบบอย่าง การนำเสนอข้อมูลประเภทนี้ที่นิยมคือ การนำเสนอข้อมูลในรูป ตาราง การนำเสนอข้อมูลในรูปแผนภาพและแผนภูมิหรือการนำเสนอข้อมูลในรูป กราฟ



ตัวอย่าง

“...จากการรายงานของสำนักงานสถิติแห่งชาติ เกี่ยวกับการลงทุนของชาวต่างชาติ ตั้งแต่เดือนมกราคม พ.ศ. 2539 ถึงเดือนธันวาคม พ.ศ. 2539 ปรากฏว่าเป็นการลงทุนจากประเทศสหรัฐอเมริกา 6,900 ล้านบาท จากประเทศเกาหลีใต้ 350 ล้านบาท จากประเทศญี่ปุ่น 530 ล้านบาท และจากประเทศสิงคโปร์ 160 ล้านบาท...”

ตารางแสดงการลงทุนของชาวต่างชาติ ระหว่างเดือนมกราคม - ธันวาคม 2539

หน่วย : ล้านบาท

ประเทศ	การลงทุน
สหรัฐอเมริกา	6,900
ญี่ปุ่น	530
เกาหลี	350
สิงคโปร์	160

ที่มา : สำนักงานสถิติแห่งชาติ

เปรียบเทียบการนำเสนอข้อมูลในลักษณะไม่เป็นแบบแผน
และในลักษณะมีแบบแผน จะเห็นว่า การนำเสนอข้อมูลทั้ง
สองแบบมีทั้งข้อดีและข้อเสียแตกต่างกัน ขึ้นอยู่กับการใช้
งาน เช่น ถ้าเป็นรายงานสรุปก็ควรจะเป็นแบบไม่มี
แบบแผน แต่ ถ้าต้องการศึกษาลักษณะของข้อมูล
วิเคราะห์เปรียบเทียบก็ควรใช้แบบมีแบบแผน