บทที่ 2

สถิติพรรณนา

จากความหมายของสถิติที่เป็นทั้งศาสตร์และศิลป์นั้นหมายความว่าการใช้สถิติเป็นเครื่องมือประกอบการตัดสินใจ และวางแผนเรื่องต่าง ๆ จะต้องมีความรู้และขั้นตอนที่เหมาะสม ดังนั้นหลังจากการเก็บรวบรวมข้อมูลตามจำนวนตัวอย่าง และวิธีการที่เหมาะสมแล้ว ขั้นต่อมาในระเบียบวิธีทางสถิติคือการนำเสนอข้อมูล หรือการวิเคราะห์ข้อมูลเบื้องต้น โดยอาศัยสถิติเชิงพรรณนาเพื่ออธิบายลักษณะ หรือบอกภาพกว้าง ๆ ของข้อมูลที่เก็บรวบรวมมาตามวัตถุประสงค์ของผู้ใช้จะทำให้เกิดความเข้าใจในข้อมูลมากขึ้น เทคนิคการวิเคราะห์ข้อมูลที่ใช้ในส่วนนี้ได้แก่การแจกแจงความถี่ การวัดแนวโน้มเข้าสู่ส่วนกลาง การวัดการกระจาย สัดส่วนเป็นต้น

เทคนิคการวิเคราะห์ข้อมูลในส่วนของสถิติพรรณนา

สถิติพรรณนาเป็นสถิติที่ว่าด้วยวิธีการวิเคราะห์ข้อมูลเบื้องต้น เพื่อให้ทราบว่าลักษณะโดยรวม หรือภาพกว้าง ๆ ของข้อมูลที่เก็บรวบรวมมาตามวัตถุประสงค์ของผู้ใช้ ซึ่งด้วยเทคนิคในการวิเคราะห์ข้อมูลแต่ละเทคนิคที่มีวัตถุประสงค์และวิธีการที่แตกต่างกัน ดังนี้

1 การแจกแจงความถี่

เทคนิคการแจกแจงความถี่ เป็นเทคนิคที่ทำการจำแนกข้อมูลตามลักษณะต่าง ๆ ที่ซ้ำ ๆ กันเพื่อให้ทราบว่าข้อมูลส่วนใหญ่มีค่าอยู่ที่ใด มีค่าที่ผิดปกติหรือไม่ และนอกจากนี้ยังเป็นการจัดเตรียมข้อมูลเพื่อความสะดวกในการวิเคราะห์ข้อมูลขั้นต่อไป

2. การวัดแนวโน้มเข้าสู่ส่วนกลาง

เทคนิคการวัดแนวโน้มเข้าสู่ส่วนกลาง เป็นเทคนิคที่ทำการหาตัวแทนของข้อมูลชุดหนึ่ง หรือที่เรียกว่าค่ากลาง เพื่อให้ทราบว่าในภาพรวมข้อมูลชุดนี้มีค่ามากหรือน้อย

3. การวัดการกระจาย

เทคนิคการวัดการกระจาย เป็นเทคนิคที่ทำการหาความแตกต่าง หรือการกระจายของข้อมูล เพื่อให้ทราบลักษณะของข้อมูลว่ามีค่าสอดคล้องกัน มีความแตกต่างกัน หรือมีการกระจายมากน้อยเพียงใด

4. สัดส่วน

เทคนิคสัดส่วน เป็นเทคนิคที่ทำการหาร้อยละของสิ่งที่สนใจเทียบกับจำนวนทั้งหมด เพื่อให้ทราบว่าในภาพรวมสิ่งที่สนใจมีสัดส่วนเป็นอย่างไร

การแจกแจงความถี่ (frequency distribution)

เทคนิคการแจกแจงความถี่ เป็นเทคนิคที่จำแนกข้อมูลตามลักษณะต่าง ๆ ที่ซ้ำ ๆ แล้วนับจำนวน หรือเรียกว่าความถี่ (frequency) จึงทำให้ทราบว่าข้อมูลส่วนใหญ่มีค่าเป็นเท่าไรและนอกจากนี้ยังเป็นการจัดเตรียมข้อมูลเพื่อความสะดวกในการวิเคราะห์ข้อมูลขั้นต่อไป ส่วนวิธีการในการแจกแจงความถี่มีหลายวิธี ในเอกสารฉบับนี้จะกล่าวถึง 3 วิธี คือ

1. การแจกแจงความถี่ด้วยแผนภาพลำต้นและใบ
2. การแจกแจงความถี่ด้วยตาราง
3. การแจกแจงความถี่ด้วยกราฟ
4. สัดส่วนของสิ่งที่สนใจ

1. การแจกแจงความถี่ด้วยแผนภาพลำต้นและใบ (Stem and Leaf)

การแจกแจงความถี่ด้วยแผนภาพลำต้นและใบ เป็นการแจกแจงความถี่โดยใช้ค่าของข้อมูลจริงทุกค่า ดังนั้นหลังการจำแนกข้อมูลแล้วยังสามารถเห็นลักษณะที่แท้จริงของข้อมูล การแจกแจงความถี่ด้วยแผนภาพลำต้นและใบมีขั้นตอน ดังนี้

ขั้นที่ 1 เรียงลำดับข้อมูลจากน้อยไปมาก

ขั้นที่ 2 แบ่งค่าของข้อมูลออกเป็น 2 ส่วนตามจำนวนหลักของตัวเลข ตัวเลขส่วนแรกอยู่ในส่วนของลำต้น (stem) และตัวเลขที่เหลืออยู่ในส่วนของใบ (leaf) เช่น ถ้าข้อมูลเป็นเลขสองหลัก หลักสิบจะอยู่ในส่วนของลำต้น และหลักหน่วยจะอยู่ในส่วนของใบ ถ้าข้อมูลเป็นเลข

สองหลักและสามหลักผสมกัน สำหรับเลขสองหลักให้เรียงลำดับเลขหลักสิบในส่วนของลำต้น สำหรับเลขสามหลักให้เรียงลำดับเลขหลักร้อย และหลักสิบ ในส่วนของ ลำต้น เป็นต้น

ขั้นที่ 3นำข้อมูลที่เรียงลำดับในข้อ 1 มาบันทึก โดยเรียงค่าตามลำต้นและใบ

ตัวอย่าง 2.1 จากการเลือกตัวอย่างยอดขายรายวัน (บาท) ของร้านอาหารแห่งหนึ่งจำนวน 25 วัน ได้ข้อมูลดังนี้

660	595	1060	500	630	899	1295	749	820	843
710	950	720	575	760	1090	770	682	1016	650
425	367	1480	945	1120					

จงแจกแจงความถี่ของข้อมูลข้างต้นด้วยแผนภาพลำต้นและใบ

วิธีทำ ค่าต่ำสุดและสูงสุดของยอดขายคือ 367 และ 1480 บาท ซึ่งเป็นเลขสามหลัก และสี่หลัก จึงเรียงลำดับเลขหลักแรกสำหรับเลขสามหลัก และเลขสองหลักแรกสำหรับเลขสี่หลักเป็นลำต้น โดยให้อยู่ในแนวตั้งฉาก จากนั้นในแต่ละลำต้นในแถวนอนจึงจะเรียงลำดับของเลขจากน้อยไปมาก เช่น เลขสามหลัก ที่มีค่าตั้งแต่ 600 ขึ้นไปมี 660,630,650 และ 682 คือเรียงเป็น 630 650 660 682 ซึ่งสามารถเขียนได้ดังนี้

6	30	50 60 82

ดังนั้นสามารถจัดข้อมูลทั้งหมด 25 ค่าได้ดังนี้

3	67				
4	25				
5	00	75	95		
6	30	50	60	82	
7	10	20	49	60	70
8	20	43	90		
9	45	50			
10	16	60	99		
11	20				
12	95				
13					
14	80				

จากการเก็บรวบรวมข้อมูลยอดขายรายวันของร้านค้าแห่งนี้จำนวน 25 วัน พบว่าโดยส่วนใหญ่มียอดขายรายวันประมาณ 600 – 800 บาท

2. การแจกแจงความถี่ด้วยตาราง (Frequency Table)

การแจกแจงความถี่วิธีนี้เป็นการจำแนกข้อมูลที่มีค่าซ้ำ ๆ กัน แล้วนับจำนวนครั้งของการเกิดซ้ำเรียกว่า ความถี่ (frequency) ในรูปของตารางที่เรียกว่าตารางแจกแจงความถี่ (frequency table) และสามารถจำแนกข้อมูลค่าต่าง ๆ ในรูปของร้อยละได้อีกด้วย ตารางแจกแจงความถี่แบ่งได้ 2 ประเภท ดังนี้

1. ตารางแจกแจงความถี่ชนิดไม่จัดข้อมูลเป็นหมวดหมู่ (ungrouped frequency table)
2. ตารางแจงความถี่ชนิดจัดข้อมูลเป็นหมวดหมู่ (grouped frequency table)

2.1 ตารางแจกแจงความถี่ชนิดไม่จัดข้อมูลเป็นหมวดหมู่ เป็นวิธีการที่เหมาะสมกับข้อมูลเชิงคุณภาพ และข้อมูลเชิงปริมาณที่มีค่าแตกต่างกันไม่มากนัก และมีบางค่าที่ซ้ำกัน มีขั้นตอนการสร้าง ดังนี้

ขั้นที่ 1 เขียนลักษณะต่าง ๆ ของข้อมูลสำหรับข้อมูลเชิงคุณภาพ หรือเขียนเรียงลำดับข้อมูลตั้งแต่ค่าต่ำสุดไปจนถึงค่าสูงสุดสำหรับข้อมูลเชิงปริมาณในคอลัมน์แรก

ขั้นที่ 2 พิจารณาแต่ละค่าของข้อมูลว่ามีจำนวนซ้ำเท่าใด โดยเขียนรอยขีดแทนจำนวนซ้ำของข้อมูลในคอลัมน์ที่ 2 คือคอลัมน์รอยขีด เพื่อความสะดวกในการนับความถี่ (การทำรอยขีดควรทำให้เป็นหมวดหมู่ เช่น ////).

ขั้นที่. นับจำนวนรอยขีดแล้วเขียนจำนวนนั้นในคอลัมน์ที่ 3 คือคอลัมน์ความถี่ ซึ่งมักจะเขียนแทนด้วยสัญลักษณ์ f ถ้าข้อมูลค่าใดไม่มีรอยขีดเลย แสดงว่ามีความถี่เป็น 0 ทั้งนี้ผลรวมของความถี่ของข้อมูลทุกค่าต้องเท่ากับจำนวนของข้อมูล

ตัวอย่าง 2.2 จงสร้างตารางแจกแจงความถี่ของข้อมูลผู้โดยสารเครื่องบินเที่ยวกรุงเทพฯ - เชียงใหม่ จำนวน 20 คน ดังนี้

คนที่	เพศ	น้ำหนักกระเป๋า (kg)
1	ช	8
2	ช	7
3	ญ	10
4	ช	9
5	ช	11
6	ช	7
7	ช	8
8	ญ	10

คนที่	เพศ	น้ำหนักกระเป๋า (kg)
9	ญ	11
10	ช	7
11	ญ	8
12	ญ	9
13	ช	11
14	ญ	11
15	ญ	10
16	ญ	8
17	ช	7
18	ช	8
19	ญ	10
20	ช	7

วิธีทำ เนื่องจากเพศเป็นข้อมูลเชิงคุณภาพ และจำแนกได้เป็น 2 ลักษณะ สามารถสร้างตารางแจกแจงความถี่ได้ ดังนี้

ตารางที่ 2.1 จำนวนผู้โดยสารเครื่องบินเที่ยวกรุงเทพ ฯ - เชียงใหม่จำแนกตามเพศ

เพศ	รอยขีด	ความถี่
ชาย	𝍸 𝍸 /	11
หญิง	𝍸 ////	9
	รวม	20

จากตารางแจกแจงความถี่หมายความว่า จากผู้โดยสารเครื่องบินเที่ยวกรุงเทพ ฯ ถึงเชียงใหม่จำนวน 20 คน เป็นเพศชาย 11 คน และเพศหญิง 9 คน

เนื่องจากน้ำหนักสัมภาระเป็นข้อมูลเชิงปริมาณมีค่าต่ำสุดเป็น 7 และค่าสูงสุดเป็น 11 ซึ่งแตกต่างกันไม่มากนัก และมีข้อมูลบางค่าที่ซ้ำกันจึงสร้างตารางแจกแจงความถี่ได้ ดังนี้

ตารางที่ 2.2 จำนวนผู้โดยสารเครื่องบินเที่ยวกรุงเทพ ฯ – เชียงใหม่จำแนกตามน้ำหนักสัมภาระ (kg)

น้ำหนักสัมภาระ(kg)	รอยขีด	ความถี่
7	𝍸	5
8	𝍸	5
9	//	2
10	////	4
11	////	4
	รวม	20

จากตารางแจกแจงความถี่หมายความว่า จากผู้โดยสารเครื่องบินเที่ยวกรุงเทพ ฯ ถึงเชียงใหม่จำนวน 20 คน ส่วนใหญ่นำสัมภาระติดตัวมาหนัก 7 - 8 กิโลกรัม

หมายเหตุ ในกรณีที่ข้อมูลเชิงปริมาณมีค่าแตกต่างกันมาก คอลัมน์แรกของตารางแจกแจงความถี่ชนิดไม่จัดข้อมูลเป็นหมวดหมู่ที่แสดงแต่ละค่าของข้อมูลก็จะมีความยาวมากเกินไป ดังนั้นตารางแจกชนิดนี้จึงเหมาะสมกับข้อมูลเชิงคุณภาพมากกว่า แต่เราก็สามารถแจกแจงความถี่ของข้อมูลเชิงปริมาณในรูปตารางแจกแจงความถี่ได้ โดยแบ่งข้อมูลเชิงปริมาณเป็นช่วง ๆ เสียก่อน ดังตารางแจกแจงความถี่ชนิดจัดข้อมูลเป็นหมวดหมู่

2.2 การแจกแจงความถี่ชนิดจัดข้อมูลเป็นหมวดหมู่ เป็นการแจกแจงความถี่ของข้อมูลเชิงปริมาณที่มีค่าแตกต่างกันมากเพื่อแก้ปัญหาจำนวนชั้นข้อมูลมากเกินไปโดยจัดข้อมูลให้เป็นช่วง ๆ หรือเป็นหมวดหมู่เสียก่อนแล้วจึงเขียนรอยขีด และนับความถี่ตามขั้นตอนของการสร้างตารางแจกแจงความถี่ที่ผ่านมา ดังตาราง 2.3

ตารางที่ 2.3 การแจกแจงน้ำหนักสัมภาระ (kg) ของผู้โดยสารเครื่องบินเที่ยวกรุงเทพ ฯ - เชียงใหม่

น้ำหนักสัมภาระ (kg)	รอยขีด	ความถี่
7-9	//	2
10-12	𝍸///	8
13-15	𝍸𝍸////	14
16-18	𝍸𝍸𝍸𝍸////	19
19-21	𝍸//	7
	รวม	50

จากตารางที่ 2.3 เป็นตารางแจกแจงความถี่น้ำหนักสัมภาระของผู้โดยสาร 50 คน ซึ่งข้อมูลถูกบันทึกในลักษณะปัดเศษให้เป็นจำนวนเต็มในหน่วยกิโลกรัม โดยน้ำหนักสัมภาระถูกแบ่งออกเป็น 5 ช่วง หรือที่เรียกว่าอันตรภาคชั้น (class interval) ดังคอลัมน์แรก คือ 7-9 10-12 13-15 16-18 และ 19-21 เรียกค่าต่ำสุดของแต่ละอันตรภาคชั้นว่าขีดจำกัดล่างของอันตรภาคชั้น (lower class limit) และเรียกค่าสูงสุดของแต่ละอันตรภาคชั้นว่าขีดจำกัดบนของอันตรภาคชั้น (upper class limit) ดังนั้นอันตรภาคชั้น 13-15 มี 13 เป็นขีดจำกัดล่าง และ 15 เป็นขีดจำกัดบน อันตรภาคชั้นนี้จะครอบคลุมค่าน้ำหนักตั้งแต่ 12.5 กิโลกรัม แต่น้อยกว่า 15.5 กิโลกรัม เรียก 12.5 ว่าขอบเขตล่างของอันตรภาคชั้น (lower class boundary) และเรียก 15.5 ว่าขอบเขตบนของอันตรภาคชั้น (upper class boundary) ซึ่งเป็นขีดจำกัดที่แท้จริงของอันตรภาคชั้น 13-15 ขอบเขตล่าง และขอบเขตบนจะมีค่าเป็นเท่าไรนั้นขึ้นอยู่กับว่าขีดจำกัดบนและขีดล่างของอันตรภาคชั้นถัดไปต่างกันอยู่เท่าไร หรือเรียกว่าผลต่างของขีดจำกัดบนและขีดล่างของอันตรภาคชั้นถัดไป เช่น ขีดจำกัดบนของอันตรภาคชั้น 10-12 ต่างจากขีดจำกัดล่างของอันตรภาคชั้น 13-15 เท่ากับ 1 หรือผลต่างของขีดจำกัดบนและขีดล่างของอันตรภาคชั้นถัดไปเท่ากับ 1 ดังนั้นขอบเขตล่างของอันตรภาคชั้น 13-15 มีค่าน้อยกว่าขีดจำกัดล่างอยู่ 0.5 และขอบเขตบนมีค่ามากกว่าขีดจำกัดบนอยู่ 0.5 ถ้าอันตรภาคชั้นต่างกันอยู่ 0.1 ขอบเขตล่างของอันตรภาคชั้นมีค่าน้อยกว่าขีดจำกัดล่างอยู่ 0.05 และขอบเขตบนมีค่ามากกว่าขีดจำกัดบนอยู่ 0.05 เป็นต้น หรือสามารถหาขอบเขตล่างและขอบเขตบนจากสูตร ดังนี้

$$\text{ขอบเขตล่าง} = \frac{\text{ขีดจำกัดล่างของอันตรภาคชั้นนั้น} + \text{ขีดจำกัดบนของอันตรภาคชั้นที่ต่ำกว่า}}{2}$$

$$\text{ขอบเขตบน} = \frac{\text{ขีดจำกัดบนของอันตรภาคชั้นนั้น} + \text{ขีดจำกัดล่างของอันตรภาคชั้นที่สูงกว่า}}{2}$$

ตารางที่ 2.4 การแจกแจงน้ำหนักสัมภาระ (kg) ของผู้โดยสารเครื่องบินเที่ยวกรุงเทพ ฯ - เชียงใหม่

ขอบเขตล่าง	น้ำหนักสัมภาระ (kg)	ขอบเขตบน	จำนวนชิ้น
6.5	7-9	9.5	2
9.5	10-12	12.5	8
12.5	13-15	15.5	14
15.5	16-18	18.5	19
18.5	19-21	21.5	7
		รวม	50

ตารางที่ 2.5 การแจกแจงความสูง (cm) ของนักเรียน

ขอบเขตล่าง	ความสูง(cm)	ขอบเขตบน	จำนวนนักเรียน
139.45	139.5-144.4	144.45	5
144.45	144.5-149.4	149.45	18
149.45	149.5-154.4	154.45	42
		รวม	65

ตารางที่ 2.6 การแจกแจงปริมาณสินค้า (ล้านตัน)

ขอบเขตล่าง	ปริมาณสินค้า (ล้านตัน)	ขอบเขตบน	จำนวนปี
	0-0.99	0.995	3
0.995	1.00-1.99	1.995	4
1.995	2.00-2.99	2.995	3
		รวม	10

จากตารางการแจกแจงความถี่ข้างต้น นอกจากจะสามารถทำความเข้าใจกับลักษณะของข้อมูลได้โดยง่ายแล้ว ตารางดังกล่าวยังประกอบข้อมูลที่จำเป็นต่อการวิเคราะห์ในขั้นต่อไปอย่างเพียงพอ แต่ในขณะเดียวกันเราจะสูญเสียรายละเอียดต่าง ๆ ของข้อมูลไป เช่น ในตารางที่ 2.4 เราไม่ทราบว่าคน 14 คนในอันตรภาคชั้น 13-15 มีสัมภาระหนักน้อยที่สุด หรือมากที่สุดกี่กิโลกรัม และไม่ทราบว่ามีกี่คนที่มีสัมภาระหนัก 13 กิโลกรัม 14 กิโลกรัม

ดังนั้นจำนวนอันตรภาคชั้นของตารางแจกแจงความถี่จึงมีความสำคัญ เพราะถ้ามีจำนวนอันตรภาคชั้นน้อยเกินไปจะทำให้สูญเสียรายละเอียดของข้อมูลมากขึ้น แต่ในขณะเดียวกันถ้ามีจำนวนอันตรภาคชั้นมากเกินไปจะทำให้ตารางมีขนาดใหญ่โดยไม่จำเป็นโดยทั่วไปแล้วตารางแจกแจงความถี่นิยมใช้จำนวนอันตรภาคชั้นตั้งแต่ 5-15 แต่อย่างไรก็ตามมิใช่กฎเกณฑ์ตายตัว เพราะต้องพิจารณาถึงค่าและจำนวนของข้อมูลแต่ละชุดเป็นสำคัญ

เนื่องจากเราไม่ทราบว่าคนจำนวน 14 คนในอันตรภาคชั้น 13-15 มีสัมภาระหนักเป็นค่าต่าง ๆ จำนวนกี่คน ดังนั้นถ้าต้องการหาตัวแทนของน้ำหนักสัมภาระในแต่ละอันตรภาคชั้นก็คือการหาจุดกึ่งกลางของอันตรภาคชั้น ดังนี้

$$\text{จุดกึ่งกลางของอันตรภ าคชั้น} = \frac{\text{ขีดจำกัดบน} + \text{ขีดจำกัดล่าง}}{2}$$

เช่นอันตรภาคชั้น 13-15 มีจุดกึ่งกลางชั้น คือ $\frac{13+15}{2}$ = 14 ซึ่งเป็นตัวแทนของน้ำหนักสัมภาระในอันตรภาคชั้น 13-15 เป็นต้น

ผลต่างระหว่างขอบเขตล่างและขอบเขตบนของแต่ละอันตรภาคชั้น คือ ความกว้างของอันตรภาคชั้น (class width) หรือแทนด้วยสัญลักษณ์ I ดังนั้น

I = (ขีดจำกัดบน - ขีดจำกัดล่าง) + ผลต่างของขีดจำกัดบนและขีดล่างของอันตรภาคชั้นถัดไป

จากตารางที่ 2.4 มีความกว้างของอันตรภาคชั้นเป็น 3 เท่ากันทุกอันตรภาคชั้น ในการสร้างตารางแจกแจงความถี่นั้นไม่จำเป็นที่จะต้องให้อันตรภาคชั้นมีความกว้างเท่ากันทุกอันตรภาชั้นก็ได้ เพราะต้องพิจารณาถึงค่าและจำนวนของข้องมูลแต่ละชุดเป็นสำคัญ

ในบางกรณีที่ข้อมูลบางค่ามีค่าน้อยหรือมากผิดปกติ บางอันตรภาคชั้นของตารางแจกแจงความถี่จะมีลักษณะเป็นอันตรภาคชั้นเปิด (open class) โดยในชั้นดังกล่าวจะระบุขีดจำกัดล่างหรือขีดจำกัดบนของอันตรภาคชั้นเพียงค่าเดียว และจะมีคำว่าน้อยกว่า หรือมากกว่ารวมอยู่ด้วย ทั้งนี้เพื่อให้มีจำนวนชั้นที่ครอบคลุมข้อมูลทั้งหมดไม่เกินไปนั่นเอง ดังตารางที่ 2.7

ตารางที่ 2.7 การแจกแจงความถี่ระดับเชาว์ปัญญาของนักเรียน

เชาว์ปัญญา	ความถี่
น้อยกว่า 80	5
80-89	10
90-99	19
100-109	23
110-119	20
120-129	16
มากกว่า 129	7
รวม	100

หมายเหตุ โดยทั่วไปทั้งกรณีที่ความกว้างของแต่ละอันตรภาคชั้นไม่เท่ากัน และมีบางอันตรภาคชั้นเป็นอันตรภาคชั้นเปิดนั้นไม่เป็นที่นิยม เนื่องจากไม่สะดวกในการพิจารณา และจะทำให้การวิเคราะห์ข้อมูลในขั้นต่อไปนั้นทำได้ยาก

2.2.1 การสร้างตารางแจกแจงความถี่ชนิดจัดข้อมูลเป็นหมวดหมู่ ในการสร้างตารางแจกแจงความถี่ประเภทนี้ต้องคำนึงถึงความเหมาะสมของจำนวนและความกว้างของอัตรภาคชั้น ซึ่งมีขั้นตอนในการสร้าง ดังนี้

ขั้นที่ 1. กำหนด หรือคำนวณจำนวนอันตรภาคชั้นจากสูตรของ Sturges

$$k = 1+3.3\log(n)$$

เมื่อ k แทนจำนวนอันตรภาคชั้น และ n แทนจำนวนข้อมูล

ขั้นที่ 2. คำนวณหาพิสัย (range) จาก

$$R = \text{ค่าสูงสุด} - \text{ค่าต่ำสุด}$$

เมื่อ R แทนพิสัย

ขั้นที่ 3. คำนวณหาความกว้างของอันตรภาคชั้นจาก

$$I = \frac{\text{พิสัย}}{\text{จำนวนอันตรภาคชั้น}}$$

เมื่อ I แทนความกว้างของอันตรภาคชั้น

ขั้นที่ 4. กำหนดขีดจำกัดล่างของอันตรภาคชั้นแรกซึ่งควรเป็นค่าที่สอดคล้องกับข้อมูล เช่นถ้าข้อมูลเป็นเลขจำนวนเต็ม ขีดจำกัดล่างก็ควรเป็นเลขจำนวนเต็มด้วยถ้าข้อมูลเป็นทศนิยม 1 ตำแหน่ง ขีดจำกัดล่างก็ควรเป็นทศนิยม 1 ตำแหน่งด้วยเป็นต้น แล้วคำนวณหาขีดจำกัดบน โดยนำขอบเขตล่างบวกด้วยความกว้างของอันตรภาคชั้นจะได้ขอบเขตบนของอันตรภาคชั้น จากนั้นเขียนอันตรภาคชั้นให้ครบตามที่ต้องการ หรือคำนวณค่าขีดจำกัดล่างของอันตรภาคชั้น จาก

$$\text{ขีดจำกัดล่าง} = \text{ค่าต่ำสุด} - \frac{Ik - R}{2}$$

เมื่อ I แทนความกว้างของอันตรภาคชั้น

R แทนพิสัย

k แทนจำนวนอันตรภาคชั้น

และคำนวณขีดจำกัดบน จาก

ขีดจำกัดบน = (ขีดจำกัดล่าง+I) - ผลต่างของขีดจำกัดบนและขีดล่างของอันตรภาคชั้นถัดไป

5. เขียนรอยขีด และนับความถี่ของแต่ละอันตรภาคชั้น

หมายเหตุ ถ้าความกว้างของอันตรภาคชั้นที่คำนวณได้เป็นทศนิยมและข้อมูลมีค่าแตกต่างกันมาก เพื่อความสะดวกในการสร้างอันตรภาคชั้นอาจปัดขึ้นเป็นจำนวนเต็มได้

ตัวอย่าง 2.3 จงสร้างตารางแจกแจงความถี่ที่มี 9 อันตรภาคชั้นของรายได้ต่อเดือน (ร้อยบาท) ของแม่ค้าในตลาดห้วยพลูจำนวน 150 คน

27	79	69	40	51	88	55	48	36	61	53	44	85	61	59
94	51	65	42	58	55	69	63	70	48	61	55	62	57	48
60	25	47	78	61	54	57	76	73	62	36	67	80	57	42
40	51	59	68	27	67	62	43	54	83	59	12	89	60	51
72	57	82	45	54	52	71	53	82	69	60	35	69	76	52
41	65	62	85	60	42	55	34	49	45	49	64	45	59	44
40	61	73	44	59	46	71	86	43	69	54	31	81	53	58
56	51	75	44	66	53	80	71	53	56	91	60	49	45	54
41	29	56	57	35	51	43	39	56	27	62	44	68	73	55
26	77	67	41	33	61	70	39	58	69	51	85	46	55	67

วิธีทำ สร้างตารางแจกแจงความถี่ตามขั้นตอน ดังนี้

ขั้นที่ 1 กำหนดจำนวนอันตรภาคชั้น เท่ากับ 9

ขั้นที่ 2 คำนวณหาค่าพิสัย จาก

$$\begin{aligned} R &= \text{ค่าสูงสุด - ค่าต่ำสุด} \\ &= 94 - 12 \\ &= 82 \end{aligned}$$

ขั้นที่ 3 คำนวณความกว้างของอันตรภาคชั้น จาก

$$\begin{aligned} I &= \frac{\text{พิสัย}}{\text{จำนวนอันตรภาคชั้น}} \\ &= \frac{82}{9} \\ &= 9.11 \end{aligned}$$

เนื่องจากข้อมูลเป็นเลขจำนวนเต็ม และมีค่าแตกต่างกันมาก เพื่อความสะดวกจึงประมาณความกว้างของอันตรภาคชั้นเป็นเลขจำนวนเต็มได้ ดังนั้นความกว้างของอันตรภาคชั้นเท่ากับ 10

ขั้นที่ 4 กำหนดขีดจำกัดล่างเป็น 10 และคำนวณหาขีดจำกัดบน ดังนี้

ขอบเขตล่าง + ความกว้างของอันตรภาคชั้น	ขอบเขตบน	ขีดจำกัดบน	อันตรภาคชั้น
9.5 + 10	19.5	19	10-19
19.5 + 10	29.5	29	20-29
29.5 + 10	39.5	39	30-39
39.5 + 10	49.5	49	40-49
49.5 + 10	59.5	59	50-59
59.5 + 10	69.5	69	60-69
69.5 + 10	79.5	79	70-79
79.5 + 10	89.5	89	80-89
89.5 + 10	99.5	99	90-99

ขั้นที่ 5 เขียนรอยขีดและนับความถี่ จะได้ตารางแจกแจงความถี่ ดังตารางที่ 2.8

ตารางที่ 2 8 จำนวนแม่ค้าในตลาดห้วยพลู จำแนกตามรายได้ช่วงต่าง ๆ

รายได้ (ร้อยบาท)	รอยขีด	จำนวน
10-19	/	1
20-29	𝍸 /	6
30-39	𝍸 ////	9
40-49	𝍸 𝍸 𝍸 𝍸 𝍸 𝍸	30
50-59	𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 //	42
60-69	𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 ///	33
70-79	𝍷 𝍷 𝍷	15
80-89	𝍷 𝍷 //	12
90-99	//	2
รวม		150

จากตารางจะเห็นว่าแม่ค้าในตลาดห้วยพลู 150 คนมีรายได้ในช่วง 50 – 59 ร้อยบาท จำนวนมากที่สุด

ตัวอย่าง 2.4 จากตัวอย่าง 2.3 จงคำนวณหาค่าจำนวนอันตรภาคชั้น ขีดจำกัดล่าง ขีดจำกัดบน และสร้างตารางแจกแจงความถี่

ขันที่ 1 คำนวณค่าจำนวนอันตรภาคชั้นจาก

$$\begin{aligned} k &= 1+3.3\log(n) \\ &= 1+3.3\log(150) \\ &= 1+3.3(2.18) \\ &= 8.194 \\ &\cong 9 \end{aligned}$$

ขั้นที่ 2 พิสัยเท่ากับ 82

ขั้นที่ 3 ความกว้างของอันตรภาคชั้นเท่ากับ 10

ขั้นที่ 4 คำนวณค่าขีดจำกัดล่างของอันตรภาคชั้นจาก

$$\begin{aligned} \text{ขีดจำกัดล่าง} &= \text{ค่าต่ำสุด} - \frac{Ik - R}{2} \\ &= 12 - \frac{(10)(9) - 82}{2} \\ &= 8 \end{aligned}$$

และคำนวณขีดจำกัดบนจาก

ขีดจำกัดบน = (ขีดจำกัดล่าง+I) – ผลต่างของขีดจำกัดบนและขีดล่างของอันตรภาคชั้นถัดไป

= (8 + 10) – 1

= 17

จะได้อัตรภาคชั้นดังนี้ 8 – 17 18 – 27 28 – 37 ... 88 – 97

ขั้นที่ 5 จะได้ตารางแจกแจงความถี่ ดังตารางที่ 2.9

ตารางที่ 2.9 จำนวนแม่ค้าในตลาดห้วยพลูจำแนตามรายได้ช่วงต่าง ๆ

รายได้ (ร้อยบาท)	จำนวน
8 – 17	1
18 – 27	5
28 – 37	8
38 – 47	26
48 – 57	40
58 – 67	33
68 – 77	21
78 – 87	12
88 – 97	4
รวม	150

ในบางครั้งข้อมูลที่เก็บรวบรวมมานั้นอาจเป็นทศนิยม และมีค่าแตกต่างกันน้อย เช่น ข้อมูลจากการวัด การชั่ง การแจกแจงความถี่ของข้อมูลลักษณะนี้จะแสดงด้วยตัวอย่าง 2.5

ตัวอย่าง 2.5 ข้อมูลจากการวัดความยาวของไม้ที่ตัดโดยการประมาณด้วยสายตามีหน่วยเป็นเซนติเมตร เป็นดังนี้ 7.5 6.3 7.7 … 9.5 จงสร้างอันตรภาคชั้นจำนวน 5 อันตรภาคชั้น

วิธีทำ 1. กำหนดจำนวนอันตรภาคชั้น เท่ากับ 5

2. พิสัย = 9.5 – 6.3 = 3.2

3. ความกว้างของอันตรภาคชั้น (I) = 3.2/5 = 0.64 และเนื่องจากข้อมูลมีค่าเป็นทศนิยมหนึ่งตำแหน่งดังนั้น ความกว้างของอันตรภาคชั้นเท่ากับ 0.7

4. กำหนดขีดจำกัดล่างเป็น 6.2 จะได้ขีดจำกัดบน และอันตรภาคชั้น ดังนี้

ความยาว
6.2 – 6.8
6.9 – 7.5
7.6 – 8.2
8.3 – 8.9
9.0 – 9.6

2.2.2 การแจกแจงความถี่สัมพัทธ์ การนำเสนอข้อมูลที่ดัดแปลงมาจากตารางแจกแจงความถี่ คือ ตารางแจกแจงความถี่สัมพัทธ์ (relative frequency table) โดยที่

$$\text{ความถี่สัมพัทธ์ของอันตรภาคชั้น} = \frac{\text{ความถี่ของอันตรภาคชั้น}}{\text{ความถี่รวม}}$$

และผลรวมของความถี่สัมพัทธ์จะเท่ากับ 1 ถ้าคูณแต่ละค่าของความถี่สัมพัทธ์ด้วย 100 เพื่อทำให้เป็นร้อยละผลรวมของค่าดังกล่าวทั้งหมดจะเท่ากับ 100% และเรียกว่าการแจกแจงร้อยละ (percentage distribution) จากตัวอย่าง 2.3 ได้ตารางแจกแจงความถี่สัมพัทธ์ ดังนี้

ตารางที่ 2.10 จำนวนและร้อยละของแม่ค้าในตลาดห้วยพลูจำแนกตามรายได้ช่วงต่าง ๆ

รายได้ (ร้อยบาท)	ความถี่	ความถี่สัมพัทธ์	ร้อยละ
10-19	1	0.0067	0.67
20-29	6	0.0400	4.00
30-39	9	0.0600	6.00
40-49	30	0.2000	20.00
50-59	42	0.2800	28.00
60-69	33	0.2200	22.00
70-79	15	0.1000	10.00
80-89	12	0.0800	8.00
90-99	2	0.0133	1.33
รวม	150	1.00	100

จากตารางที่ 2.10 บอกให้ทราบว่าจากแม่ค้าในตลาดห้วยพลูจำนวน 150 คน มีรายได้ 50-59 ร้อยบาทจำนวน 42 คน หรือคิดเป็น 28%

2.2.3 การแจกแจงความถี่สะสม ในบางกรณีเราไม่ได้มุ่งความสนใจเพียงแต่ความถี่ของแต่ละอันตรภาคชั้น แต่สนใจความถี่ทั้งหมดที่มีค่ามากกว่า หรือน้อยกว่าของอันตรภาคชั้นใดชั้นหนึ่ง เราเรียกความถี่ที่ได้นี้ว่าความถี่สะสม (cumulative frequency) และเรียกตารางที่ได้นี้ว่า ตารางแจกแจงความถี่สะสม (cumulative frequency table) การแจกแจงความถี่สะสม มี 2 วิธี คือ

1) การแจกแจงความในถี่สะสมชนิดน้อยกว่า (less-than cumulative distribution) เป็นตารางที่แสดงความถี่สะสมของทุกอันตรภาคชั้นที่มีค่าน้อยกว่าขอบเขตบนของแต่ละอันตรภาคชั้น

ตารางที่ 2.11 จำนวนแม่ค้าในตลาดห้วยพลูจำแนกตามรายได้ช่วงต่าง ๆ

รายได้ (ร้อยบาท)	ความถี่	คะแนน	ความถี่สะสม
10-19	1	น้อยกว่า 19.5	1
20-29	6	น้อยกว่า 29.5	7
30-39	9	น้อยกว่า 39.5	16
40-49	30	น้อยกว่า 49.5	46
50-59	42	น้อยกว่า 59.5	88
60-69	33	น้อยกว่า 69.5	121
70-79	15	น้อยกว่า 79.5	136
80-89	12	น้อยกว่า 89.5	148
90-99	2	น้อยกว่า 99.5	150

จากแม่ค้าในตลาดห้วยพลูจำนวน 150 คนมีแม่ค้าที่มีรายได้น้อยกว่า 59.5 ร้อยบาท จำนวน 88 คน

2) การแจกแจงความถี่สะสมชนิดมากกว่า (more-than cumulative distribution) เป็นตารางแสดงความถี่สะสมของทุกอันตรภาคชั้นที่มีค่ามากกว่าขอบเขตล่างของแต่ละอันตรภาคชั้น

ตารางที่ 2.12 จำนวนแม่ค้าในตลาดห้วยพลูจำแนกตามรายได้ช่วงต่าง ๆ

รายได้ (ร้อยบาท)	ความถี่	คะแนน	ความถี่สะสม
10-19	1	มากกว่า 9.5	150
20-29	6	มากกว่า 19.5	149
30-39	9	มากกว่า 29.5	143
40-49	30	มากกว่า 39.5	134
50-59	42	มากกว่า 49.5	104
60-69	33	มากกว่า 59.5	62
70-79	15	มากกว่า 69.5	29
80-89	12	มากกว่า 79.5	14
90-99	2	มากกว่า 90.5	2

จากแม่ค้าในตลาดห้วยพลูจำนวน 150 คนมีแม่ค้าที่มีรายได้มากกว่า 49.5 ร้อยบาท จำนวน 104 คน

ในทำนองเดียวกันเราสามารถสร้างตารางแจกแจงความถี่สัมพัทธ์สะสมได้ ดังนี้

ตารางที่ 2.13 การแจกแจงความถี่สัมพัทธ์สะสมรายได้ของแม่ค้า

รายได้ (ร้อยบาท)	ความถี่	ความถี่สัมพัทธ์	ความถี่สัมพัทธ์สะสม
10-19	1	0.0067	0.0067
20-29	6	0.0400	0.0467
30-39	9	0.0600	0.1067
40-49	30	0.2000	0.3067
50-59	42	0.2800	0.5867
60-69	33	0.2200	0.8067
70-79	15	0.1000	0.9067
80-89	12	0.0800	0.9867
90-99	2	0.0133	1.0000
	รวม	1.00	

จากแม่ค้าในตลาดห้วยพลูจำนวน 150 คนมีแม่ค้าที่มีรายได้น้อยกว่า 59.5 ร้อยบาท จำนวน 88 คน คิดเป็น 58.67%

3 การแจกแจงความถี่ด้วยกราฟ (graphing frequency distribution)

การแจกแจงความถี่ด้วยกราฟเป็นการแจกแจงความถี่ต่อจากตารางแจกแจงความถี่ในลักษณะที่เป็นรูปกราฟ ซึ่งทำได้ 3 วิธี คือ

1. ฮิสโตแกรม (histogram)
2. รูปหลายเหลี่ยมความถี่ (frequency polygon)
3. โค้งความถี่ (frequency curve)

3.1 ฮิสโตแกรม การนำเสนอข้อมูลวิธีนี้มีลักษณะเป็นการแสดงความถี่ของข้อมูลด้วยพื้นที่ของรูปสี่เหลี่ยมผืนผ้าที่สร้างขึ้นติดต่อกันบนแกนนอน ดังนั้นผลรวมของพื้นที่ของรูปสี่เหลี่ยมผืนผ้าทุกแท่งจะเท่ากับจำนวนข้อมูลทั้งหมดที่นำมาแจกแจง วิธีการสร้างฮิสโตแกรมในกรณีจัดข้อมูลเป็นหมวดหมู่มีขั้นตอน ดังนี้

ขั้นที่ 1 ลากแกน X และแกน Y ตัดกันเป็นมุมฉากที่จุด 0 โดยแกน X แสดงค่าของข้อมูลในแต่ละอันตรภาคชั้น และแกน Y แสดงความถี่ในแต่ละอันตรภาคชั้น เฉพาะในแกน X สามารถระบุค่าได้สองแบบ คือแบบแรกระบุค่าขอบเขตล่างและขอบเขตบนของแต่ละอันตรภาคชั้น แบบที่สองระบุจุดกึ่งกลางของแต่ละอันตรภาคชั้น

ขั้นที่ 2 บนแกน X เขียนแท่งสี่เหลี่ยมผืนผ้าให้เรียงชิดติดกัน โดยให้จุดกึ่งกลางของแต่ละแท่งคือจุดกึ่งกลางของแต่ละอันตรภาคชั้น ดังนั้นจำนวนแท่งต้องเท่ากับจำนวนอันตรภาคชั้นเสมอ

ขั้นที่ 3 ในกรณีที่ความกว้างของแต่ละอันตรภาคชั้นเท่ากันสี่เหลี่ยมผืนผ้าแต่ละแท่งมีความกว้างเท่ากัน และให้ความถี่ของแต่ละอันตรภาคชั้นเป็นความสูงของแต่ละแท่งได้เลย แต่ในกรณีที่ความกว้างของแต่ละอันตรภาคชั้นไม่เท่ากัน จะต้องหาความสูงของสี่เหลี่ยมผืนผ้าจาก

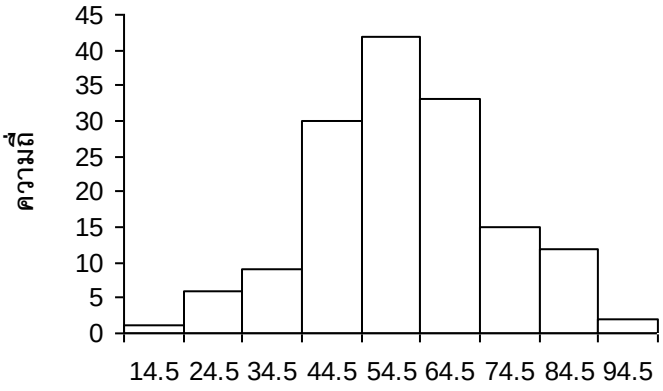
$$\text{ความสูง} = \frac{\text{ความถี่ของอันตรภาคชั้นนั้น}}{\text{สัดส่วนความกวางของอนตรภาคชนนน}}$$

ตัวอย่าง 2.6 จากตารางที่ 2.8 จงสร้างฮิสโตแกรมความถี่

รายได้ (ร้อยบาท)	จุดกึ่งกลางของอันตรภาคชั้น	ความถี่
10-19	14.5	1
20-29	24.5	6
30-39	34.5	9
40-49	44.5	30
50-59	54.5	42
60-69	64.5	33
70-79	74.5	15
80-89	84.5	12
90-99	94.5	2
		150

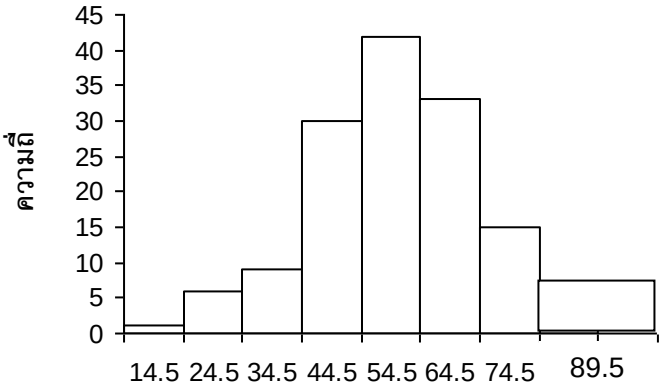
จากตารางข้างต้นจะเห็นว่าความกว้างของอันตรภาคชั้นทุกอันตรภาคชั้นเท่ากัน คือ 10 ดังนั้นความกว้างของแท่งสี่เหลี่ยมผืนผ้าจึงเท่ากับ 1 หน่วยเท่ากันทุกแท่ง ดังนั้นความสูงของแท่งสี่เหลี่ยมผืนผ้าแต่ละแท่งจึงสูงเท่ากับความถี่ของแต่อันตรภาคชั้น ดังนี้

อิสโตแกรมความถี่ของรายได้



แต่ถ้ารวมอันตรภาคชั้น 80-89 และ 90-99 ไว้ด้วยกัน จะทำให้ความกว้างของอันตรภาคชั้น 80-99 เท่ากับ 20 ดังนั้นความกว้างของแท่งสี่เหลี่ยมสุดท้ายจะเท่ากับ 2 หน่วย ดังนั้นความสูงของแท่งสี่เหลี่ยมเท่ากับ 14/2 = 7 จะได้ ฮิสโตแกรม ดังนี้

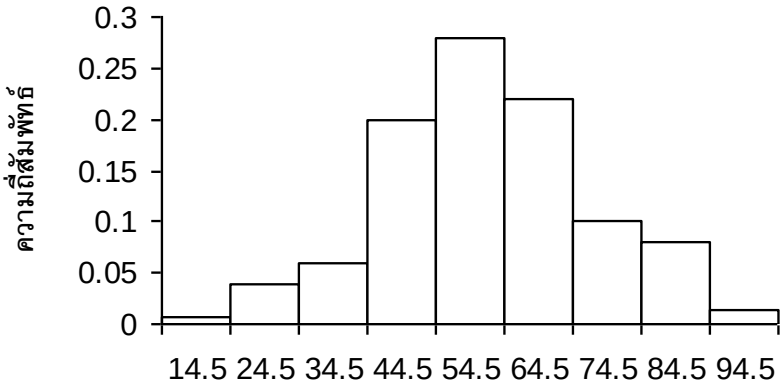
อิสโตแกรมความถี่ของรายได้



จากรูปฮิสโตแกรมที่แกน X แสดงค่าข้อมูล (หรืออันตรภาคชั้น) และแกน Y แสดงความถี่เรียกว่าฮิสโตแกรมความถี่ ซึ่งมีพื้นที่ทุกแท่งรวมกันเท่ากับความถี่รวม

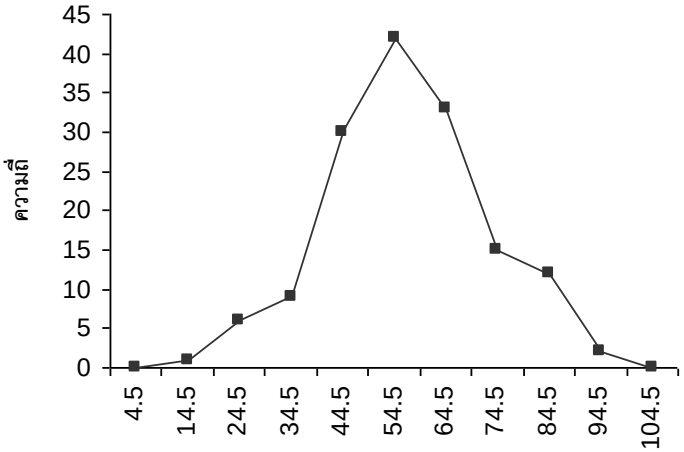
ส่วนฮิสโตแกรมที่แกน X แสดงค่าข้อมูล (หรืออันตรภาคชั้น) และแกน Y แสดงความถี่สัมพัทธ์ หรือร้อยละเรียกว่าฮิสโตแกรมความถี่สัมพัทธ์ (relative frequency histogram) หรือฮิสโตแกรมร้อยละ (percentage histogram) ซึ่งมีพื้นที่ทุกแท่งรวมกันเท่ากับ 1 หรือ 100% ดังนี้

อิสโตแกรมความถี่สัมพัทธ์ของรายได้



3.2 รูปหลายเหลี่ยมความถี่ คือกราฟเส้นที่ได้จากการลากเส้นตรงเชื่อมจุดกึ่งกลางปลายยอดทุกแท่งของฮิสโตแกรม ซึ่งเป็นผลมาจากรูปฮิสโตแกรม ดังนี้

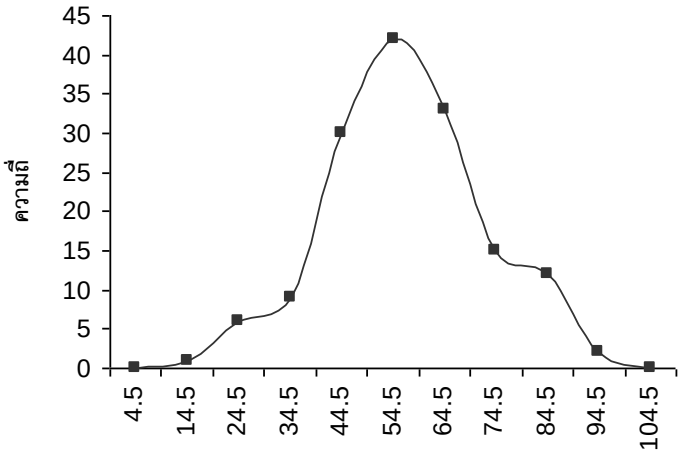
รูปหลายเหลี่ยมความถี่ของรายได้



จากรูปสังเกตว่ามีการขยายค่ากึ่งกลางอันตรภาคชั้นออกไปทั้งสองด้าน คือ ด้านที่ต่ำกว่าอันตรภาคชั้นที่หนึ่งและด้านที่สูงกว่าอันตรภาคชั้นสุดท้าย เพื่อให้เส้นกราฟสัมผัสกับแกน X จะทำให้พื้นที่ระหว่างรูปหลายเหลี่ยมความถี่และแกน X มีค่าเท่ากับผลรวมพื้นที่ทุกแท่งของฮิสโตแกรม และเรียกรูปหลายเหลี่ยมความถี่ที่แกน X แสดงค่าข้อมูล (หรืออันตรภาคชั้น) และแกน Y แสดงความถี่สัมพัทธ์ว่ารูปหลายเหลี่ยมความถี่สัมพัทธ์ ซึ่งมีพื้นที่ทุกแท่งรวมกันเท่ากับ 1

3.3 โค้งความถี่ คือเส้นโค้งที่ได้จากการปรับรูปหลายเหลี่ยมความถี่ให้เรียบโดยพื้นที่ภายใต้เส้นโค้งความถี่จะเท่ากับพื้นที่ของรูปหลายเหลี่ยมความถี่ และเรียกเส้นโค้งที่เกิดจากการปรับรูปหลายเหลี่ยมความถี่สัมพัทธ์ให้เรียบว่าโค้งความถี่สัมพัทธ์ (relative frequency curve) โดยพื้นที่ใต้โค้งความถี่สัมพัทธ์จะเท่ากับ 1

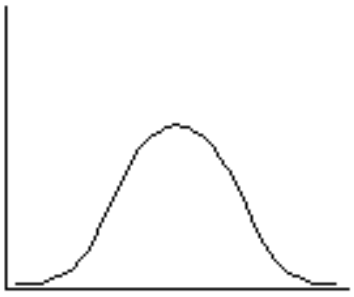
รูปโค้งความถี่ของรายได้



4. ชนิดของโค้งความถี่

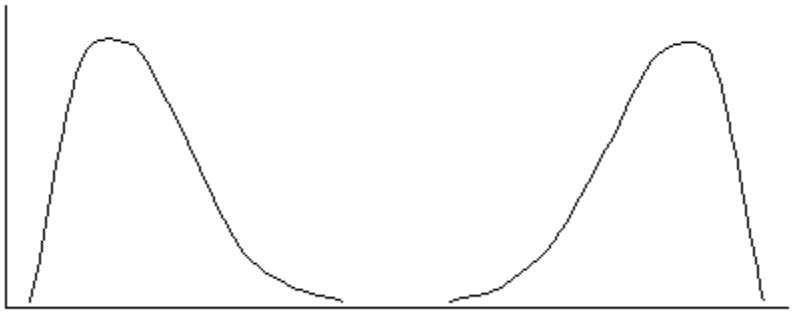
จากโค้งความถี่ หรือโค้งความถี่สัมพัทธ์ บอกให้ทราบถึงลักษณะการแจกแจงความถี่ของข้อมูล สามารถแบ่งตามรูปร่างของโค้งความถี่ได้ 6 ชนิด คือ

4.1 รูปสมมาตรหรือระฆัง (symmetrical หรือ bell-shaped) ได้แก่โค้งที่มียอดแสดงความถี่สูงสุดอยู่ตรงกลางพอดี และถ้าพับครึ่งโค้ง ณ จุดดังกล่าวปลายโค้งทั้งสองด้านจะทับกันพอดี ซึ่งหมายความว่าทั้งสองด้านมีความถี่เท่ากัน ลักษณะโค้งแบบนี้เรียกอีกอย่างว่าโค้งปกติ (normal curve) ซึ่งเป็นโค้งที่มีความสำคัญเป็นอย่างมากในวิชาสถิติ



รูปโค้งปกติ (normal curve)

4.2 รูปไม่สมมาตรหรือเบ้ (moderately asymmetrical หรือ skewed) ได้แก่โค้งที่มีปลายโค้งด้านใดด้านหนึ่งลาดยาวกว่าอีกด้านหนึ่ง ถ้าปลายด้านหนึ่งลาดยาวไปทางด้านขวาจะเรียกว่าโค้งเบ้ขวา (skewed to right) หมายความว่าข้อมูลที่มีค่าน้อยมีความถี่มากกว่าข้อมูลที่มีค่ามาก ถ้าปลายด้านหนึ่งลาดยาวไปทางด้านซ้ายจะเรียกว่าโค้งเบ้ซ้าย(skewed to left) หมายความว่าข้อมูลที่มีค่ามากมีความถี่มากกว่าข้อมูลที่มีค่าน้อย



รูปโค้งเบ้ขวา (skewed to right) รูปโค้งเบ้ซ้าย(skewed to left)

หมายเหตุ จะเห็นได้ว่าเทคนิคการแจกแจงความถี่เป็นเทคนิคที่เหมาะสมกับข้อมูลเชิงคุณภาพ แต่ถ้าเป็นข้อมูลเชิงปริมาณจะเหมาะสมกับข้อมูลเชิงปริมาณที่มีค่าซ้ำ ๆ กัน

การวัดแนวโน้มเข้าสู่ส่วนกลาง (Measures of Central Tendency)

ในบางกรณีจากข้อมูลที่เก็บรวบรวมมาได้นั้น ผู้วิเคราะห์อาจต้องการทราบตัวแทนของข้อมูลชุดนั้น เรียกตัวแทนนั้นว่าค่ากลาง ค่ากลางที่ดีควรมีค่าใกล้เคียงกับค่าส่วนใหญ่ของข้อมูล ค่ากลางจะช่วยให้เข้าใจภาพรวมของข้อมูลชุดนั้นได้เร็วกว่าการพิจารณาจากตารางแจกแจงความถี่หรือกราฟแจกแจงความถี่ ที่ต้องพิจารณาทั้งตารางหรือรูป

การคำนวณหาค่ากลาง หรือที่เรียกว่าการวัดแนวโน้มเข้าสู่ส่วนกลางของข้อมูลมีวิธีการคำนวณหลายวิธี ทั้งนี้ขึ้นอยู่กับลักษณะข้อมูล และวัตถุประสงค์ของการนำไปใช้

1. ค่าเฉลี่ยเลขคณิต (arithmetic mean)

ค่าเฉลี่ยเลขคณิตหรือเรียกว่าค่าเฉลี่ย ซึ่งแทนด้วยสัญลักษณ์ μ อ่านว่ามิว แทนค่าเฉลี่ยที่คิดจากข้อมูลของประชากร เป็นพารามิเตอร์ที่ไม่ทราบค่า ดังนั้นในทางปฏิบัติจึงหาได้เพียงค่าเฉลี่ยที่คิดจากข้อมูลของตัวอย่าง หรือค่าสถิติ ซึ่งแทนด้วยสัญลักษณ์ $\bar{x}$ อ่านว่าเอ็กซ์-บาร์ ในเอกสารฉบับนี้จะใช้สัญลักษณ์ $\bar{x}$ วิธีการหาค่าเฉลี่ยสามารถทำได้ดังนี้

กำหนดให้ข้อมูลดิบ n ตัวแทนด้วยสัญลักษณ์ $x_1, x_2, x_3, \ldots, x_n$ แล้ว $\bar{x} = \frac{x_1 + x_2 + x_3 + \ldots + x_n}{n}$ เขียนเป็นสูตรได้ดังนี้

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

เมื่อ i = 1,2,3,...,n และ n แทนจำนวนตัวอย่าง

ตัวอย่าง 2.7 จากการสำรวจราคาน้ำมันเบนซินต่อลิตรในเดือน ธันวาคม-มีนาคม 2554 ของปั๊มน้ำมันแห่งหนึ่ง ได้ข้อมูลดังนี้ 11.30 12.10 13.45 12.40 บาท จงหาราคาน้ำมันเบนซินเฉลี่ยใน 4 เดือน

$$x_1 = 11.30 \quad x_2 = 12.10 \quad x_3 = 13.45 \quad x_4 = 12.40$$

วิธีทำ

$$\begin{aligned}\bar{x} &= \frac{\sum_{i=1}^{4} x_i}{4} \\ &= \frac{11.30 + 12.10 + 13.45 + 12.40}{4} \\ &= 12.31\end{aligned}$$

ราคาน้ำมันเบนซินเฉลี่ยต่อลิตรใน 4 เดือนราคา 12.31 บาท

หมายเหตุ 1. ค่าเฉลี่ยนี้ใช้กับข้อมูลเชิงปริมาณ

2. ค่าเฉลี่ยเหมาะสมกับข้อมูลที่มีค่าสอดคล้องกัน นั่นคือไม่มีค่าผิดปกติ (ค่าที่มากหรือน้อยจนเกินไป) หรือมีโค้งความถี่เป็นโค้งสมมาตร

3. ค่าเฉลี่ยเหมาะสมกับข้อมูลที่มีจำนวนมาก ถ้าตัดค่าที่มากหรือน้อยเกินไปแล้วจะไม่มีผลต่อค่าเฉลี่ย

2. มัธยฐาน (median)

มัธยฐานซึ่งแทนด้วยสัญลักษณ์ Me หมายถึงค่ากลางที่มีตำแหน่งอยู่ตรงกลางของข้อมูลที่จัดเรียงจากน้อยไปหามาก หรือมากไปหาน้อยเรียบร้อยแล้ว ค่ามัธยฐานจึงแบ่งข้อมูล

ออกเป็น 2 กลุ่มเท่า ๆ กัน คือกลุ่มที่มีค่าน้อยกว่าค่ามัธยฐาน และกลุ่มที่มีค่ามากกว่าค่ามัธยฐาน การหาค่ากลางโดยวิธีนี้มีขั้นตอน ดังนี้

ขั้นที่ 1 เรียงลำดับข้อมูลจากมากไปหาน้อย หรือน้อยไปหามาก

ขั้นที่ 2 หาตำแหน่งตรงกลางของข้อมูลจาก $\frac{n+1}{2}$

ขั้นที่ 3 หาค่ามัธยฐาน โดยที่

กรณี n เป็นเลขคี่ เช่น x_1, x_2, x_3, x_4, x_5 ตำแหน่งตรงกลางของข้อมูล คือ $\frac{5+1}{2} = 3$ ดังนั้นค่ามัธยฐาน = x_3

กรณี n เป็นเลขคู่ เช่น x_1, x_2, x_3, x_4 ตำแหน่งตรงกลางของข้อมูล คือ $\frac{4+1}{2} = 2.5$ ดังนั้นค่ามัธยฐาน = $\frac{x_2 + x_3}{2}$

ตัวอย่าง 2.8 จากการสำรวจคุณภาพของสินค้าชนิดหนึ่งจากผู้บริโภค 10 คน กำหนดให้คะแนนเต็ม 5 คะแนน โดย 0 หมายถึงคุณภาพไม่ดี 5 หมายถึงคุณภาพดีมาก ตามลำดับ ได้ผลดังนี้ 1 4 3 5 1 2 2 4 5 4 จงหาค่ากลางของข้อมูลชุดนี้โดยใช้ค่ามัธยฐาน

วิธีทำ

1. เรียงลำดับข้อมูลจากน้อยไปหามาก จะได้ 1 1 2 2 3 4 4 4 5 5
2. ตำแหน่งตรงกลางของข้อมูล คือ $\frac{10+1}{2} = \frac{11}{2} = 5.5$
3. ค่ามัธยฐาน คือ $\frac{3+4}{2} = \frac{7}{2} = 3.5$

ค่ามัธยฐานเท่ากับ 3.5 หมายความว่าจากการสำรวจคุณภาพของสินค้าชนิดหนึ่งจากผู้บริโภค 10 คน ผู้บริโภคคิดว่าคุณภาพของสินค้าชนิดนี้อยู่ในระดับคะแนน 3.5 คะแนน

หมายเหตุ 1. มัธยฐานเหมาะสำหรับข้อมูลเชิงปริมาณ หรือข้อมูลที่มีระดับการวัดเรียงอันดับขึ้นไป เพราะเป็นวิธีใช้การจัดอันดับ ส่วนข้อมูลเชิงคุณภาพหรือข้อมูลที่มีระดับการวัดแบ่งกลุ่มไม่ควรใช้มัธยฐานเพราะการจัดอันดับข้อมูลไม่สามารถทำได้

2. มัธยฐานเหมาะกับข้อมูลที่มีค่าผิดปกติ เพราะมัธยฐานเป็นค่ากลางที่มีค่าอยู่ ณ ตำแหน่งตรงกลาง ดังนั้นค่าผิดปกติจึงไม่มีผลต่อค่ามัธยฐาน

3. ฐานนิยม (mode)

ฐานนิยมซึ่งแทนด้วยสัญลักษณ์ Mo หมายถึงค่ากลางที่ได้จากข้อมูลที่มีความถี่สูงสุด ปกติมักใช้เกี่ยวกับการสำรวจ เช่น การสำรวจความนิยมทางการเมือง การสำรวจอาชีพ การสำรวจความนิยมของการใช้บริการรถประจำทาง เป็นต้น ดังนั้นค่ากลางหรือข้อมูลนี้จึงออกมาในรูปความถี่

ตัวอย่าง 2.9 จากการสำรวจของบริษัทแห่งหนึ่งเกี่ยวกับการเลือกอาชีพหลังจบการศึกษาปริญญาตรีของนักศึกษา 1700 คนโดยบริษัทแยกอาชีพออกเป็น 7 ประเภท คือ A B C D E F และ G ปรากฏผลดังนี้ B F D E C F F G A F B E C D D F G … F จงหาฐานนิยมของข้อมูลชุดนี้

วิธีทำ อาชีพ A มีความถี่ 100
อาชีพ B มีความถี่ 200
อาชีพ C มีความถี่ 200
อาชีพ D มีความถี่ 300
อาชีพ E มีความถี่ 200
อาชีพ F มีความถี่ 500
อาชีพ G มีความถี่ 200

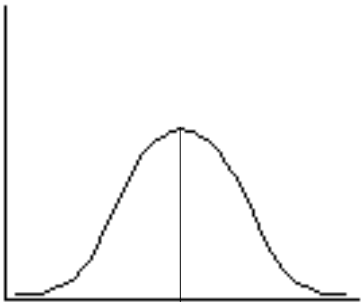
จากผลการแจกแจงความถี่ พบว่าอาชีพ F มีความถี่มากที่สุด ดังนั้นค่ากลางฐานนิยมของข้อมูลชุดนี้ อาชีพ F หมายความว่าการเลือกอาชีพหลังจบการศึกษาปริญญาตรีของนักศึกษาส่วนใหญ่เลือกอาชีพ F

หมายเหตุ ฐานนิยมสามารถใช้ได้กับข้อมูลเชิงคุณภาพ และเชิงปริมาณ แต่เหมาะสมกับข้อมูลเชิงคุณภาพมากกว่า เพราะเป็นการละทิ้งข้อสนเทศบางอย่างที่มีอยู่ในข้อมูลไป

การหาค่ากลางของข้อมูลชุดเดียวกันโดยวิธีต่าง ๆ อาจได้ค่ากลางแต่ละวิธีไม่เท่ากัน ขึ้นอยู่กับการแจกแจงของข้อมูลที่นำมาวิเคราะห์ ว่ามีการแจกแจงแบบปกติหรือไม่ นั่นหมายความว่าถ้าทราบว่าข้อมูลมีการแจกแจงแบบใด จะสามารถเลือกวิธีการหาค่ากลางได้เหมาะสม โดยพิจารณาจากความสัมพันธ์ของค่าเฉลี่ย มัธยฐาน และฐานนิยม กับโค้งการแจกแจงของข้อมูล ดังนี้

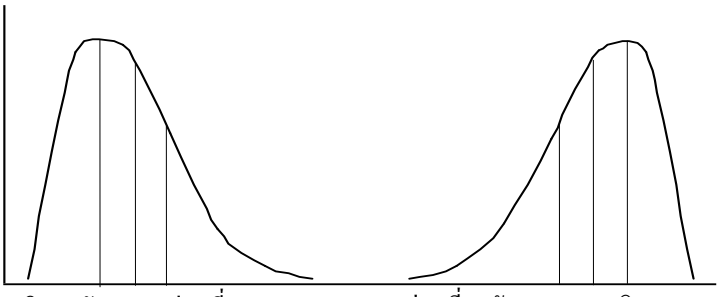
4. ความสัมพันธ์ระหว่างค่ากลางกับโค้งการแจกแจงของข้อมูล

ถ้าลักษณะของข้อมูลที่นำมาวิเคราะห์มีการแจกแจงสมมาตร หรือโค้งความถี่เป็นรูปสมมาตร (โค้งปกติ) ค่าเฉลี่ย ค่ามัธยฐาน และค่าฐานนิยมจะมีค่าเท่ากัน ดังนี้



ค่าเฉลี่ย
มัธยฐาน
ฐานนิยม

ถ้าลักษณะของข้อมูลที่นำมาวิเคราะห์มีการแจกแจงแบบเบ้ขวา หรือเบ้ซ้ายค่าเฉลี่ย ค่ามัธยฐาน และค่าฐานนิยมจะมีค่า ดังนี้



ฐานนิยม<มัธยฐาน<ค่าเฉลี่ย ค่าเฉลี่ย<มัธยฐาน<ฐานนิยม

ถ้าการแจกแจงของข้อมูลมีลักษณะเป็น โค้งเบ้ขวา แสดงว่าข้อมูลส่วนใหญ่มีค่าน้อย แสดงว่ามีข้อมูลบางค่าที่มีค่ามากเกินไปจึงมีผลกระทบทำให้ค่าเฉลี่ยมีค่ามากกว่าที่ควรจะเป็น และค่าฐานนิยมจะอยู่ ณ ตำแหน่งที่โค้งสูงที่สุด แต่ไม่มีผลกระทบต่อค่ามัธยฐานจึงทำให้มัธยฐานมีค่าน้อยกว่าค่าเฉลี่ย แต่มากกว่าฐานนิยม ดังนั้นกรณีโค้งเบ้ขวาจะได้ว่าค่า $Mo < Me < \bar{x}$ และในลักษณะตรงกันข้ามกรณีการแจกแจงของข้อมูลมีลักษณะเป็นโค้งเบ้ซ้ายจะได้ว่าค่า $\bar{x} < Me < Mo$

ในทางสถิติข้อมูลที่ดี และเหมาะสมที่จะนำไปวิเคราะห์ข้อมูลต่อไปในส่วนของสถิติอ้างอิงจะต้องมีลักษณะการแจกแจงแบบปกติ ไม่เบ้ซ้าย หรือขวา ซึ่งสามารถพิจารณาได้จากลักษณะความเบ้ของข้อมูล ดังนี้

5. ความเบ้ (Skewness)

ความเบ้ คือระดับความเอนเอียงหรือความไม่สมมาตรของการแจกแจงของข้อมูล โดยพิจารณาได้จากโค้งความถี่ หรือวัดความเบ้ (measure of skewness) จากค่าสัมประสิทธิ์ความเบ้ ซึ่งมีวิธีการคำนวณหลายวิธี ในเอกสารฉบับนี้จะใช้ MS EXCEL ในการคำนวณ ความหมายของค่าสัมประสิทธิ์ความเบ้ที่คำนวณได้เป็นดังนี้

สัมประสิทธิ์ความเบ้ = 0 แสดงว่าข้อมูลมีการแจกแจงแบบสมมาตร(Symmetrical distribution)

สัมประสิทธิ์ความเบ้ < 0 แสดงว่าข้อมูลมีการแจกแจงแบบเบ้ซ้าย (Negatively skewed distribution)

สัมประสิทธิ์ความเบ้ > 0 แสดงว่าข้อมูลมีการแจกแจงแบบเบ้ขวา (positive skewed distribution)

การวัดการกระจาย (measures of dispersion)

ในข้อมูลแต่ละชุด เมื่อพิจารณาตัวแทนของข้อมูลจากค่ากลางของข้อมูลแล้ว กรณีที่ข้อมูลแต่ละชุดมีค่ากลางเท่ากัน อาจทำให้เข้าใจว่าข้อมูลแต่ละชุดนั้นมีลักษณะที่คล้ายกัน ซึ่งในความเป็นจริงอาจไม่เป็นเช่นนั้น ดังนั้นการวิเคราะห์ข้อมูลยังมีสิ่งที่น่าสนใจต่อไปว่าข้อมูลในชุดนั้น ๆ ค่าแตกต่างกันมากน้อยแค่ไหน ลักษณะเช่นนี้เรียกว่า การวัดการกระจาย ถ้าข้อมูลทุกตัวมีค่าเท่ากันหมดเรียกว่าข้อมูลชุดนั้นมีค่าไม่แตกต่างกัน หรือไม่มีการกระจาย แต่ถ้าข้อมูลชุดนั้นมีค่าไม่เท่ากันถือว่าข้อมูลชุดนั้นมีการกระจายเกิดขึ้น หรือมีความแตกต่างระหว่างข้อมูล ขนาดการกระจายของข้อมูลขึ้นอยู่กับความแตกต่างภายในข้อมูลชุดนั้น ถ้าข้อมูลแตกต่างกันมากถือว่ามีการกระจายมาก ถ้าข้อมูลแตกต่างกันน้อยถือว่ามีการกระจายน้อย ดังตัวอย่าง 2.10

ตัวอย่าง 2.10 ในการคัดเลือกพนักงาน 1 ทีมเพื่อรับรางวัลทีมงานที่มีประสิทธิภาพในการขายดีเด่น โดยพิจารณาจากยอดขาย (หมื่นบาท) จากพนักงานแต่ละทีม ดังนี้

ทีมงาน	ยอดขาย (หมื่นบาท)							ค่าเฉลี่ย
A	55	55	55	55	55	55	55	55
B	47	51	54	55	56	59	63	55
C	39	47	53	55	57	63	71	55

จากยอดขายของพนักงานทั้ง 3 ทีมนี้ ถ้าพิจารณาที่ค่าเฉลี่ยจะพบว่าพนักงานทั้ง 3 ทีม มียอดขายเฉลี่ยเท่ากับ 55 หมื่นบาท หมายความว่าทั้ง 3 ทีมมีความสามารถในการขายเหมือนกัน แต่เมื่อพิจารณายอดขายของพนักงานแต่ละคนในแต่ละทีมจะเห็นว่า

ทีม A พนักงานมียอดขายเท่ากันทุกคน นั่นคือไม่มีการกระจายของข้อมูล หมายความว่าพนักงานในทีม A มีความสามารถในการขายเหมือนกันทุกคน

ทีม B พนักงานแต่ละคนมียอดขายแตกต่างกันเล็กน้อย นั่นคือมีการกระจายของข้อมูลมากกว่าทีม A หมายความว่าพนักงานในทีม B มีความสามารถในการขายต่างกันเล็กน้อย

ทีม C พนักงานแต่ละคนมียอดขายแตกต่างกันมาก นั่นคือมีการกระจายของข้อมูลมากกว่าทีม A และทีม B หมายความว่าพนักงานในทีม C มีความสามารถต่างกันมาก

จะเห็นว่าในภาพรวมของพนักงานทั้ง 3 ทีมมียอดขายเฉลี่ยเท่ากัน แต่เมื่อพิจารณาความสามารถของพนักงานแต่ละคนในทีม พนักงานทีม A มีความสามารถเท่ากัน นั่นหมายถึงพนักงานทีมนี้ทำงานได้มีประสิทธิภาพไม่มีใครเอาเปรียบเทียบใครสามารถทำงานเป็นทีมได้เป็นอย่างดี ดังนั้นควรให้รางวัลทีมงานที่มีประสิทธิภาพในการขายดีเด่นแก่ทีม A

วิธีการวัดการกระจายของข้อมูลนั้นมีอยู่หลายวิธี ในเอกสารฉบับนี้จะกล่าวถึงวิธี พิสัย ความแปรปรวน และสัมประสิทธิ์ความแปรผัน

1. พิสัย (range)

พิสัยซึ่งแทนด้วยสัญลักษณ์ R เป็นการวัดการกระจายของข้อมูลจากความแตกต่างระหว่างค่าสูงสุดกับค่าต่ำสุด ดังนั้น

R = ค่าสูงสุด - ค่าต่ำสุด

จากตัวอย่าง 2.10 พิสัยของยอดขายของพนักงานทีม A = 55 - 55 = 0
พิสัยของยอดขายของพนักงานทีม B = 63 - 47 = 16
พิสัยของยอดขายของพนักงานทีม C = 71 - 39 = 32

จะเห็นว่าค่าพิสัยของยอดขายของพนักงานทีม C มีค่ามากที่สุด และตามด้วยค่าพิสัยของยอดขายของพนักงานทีม B และ ทีมA ตามลำดับ หมายความว่ายอดขายของพนักงานทีม C แตกต่างกันมากกว่า ทีม B และทีม A

จะเห็นว่าการวัดการกระจายโดยใช้วิธีพิสัยนั้นเป็นวิธีที่ง่ายและรวดเร็ว แต่เป็นค่าวัดการกระจายแบบหยาบ ๆ เนื่องพิจารณาจากค่าเพียง 2 ค่าเท่านั้น คือค่าสูงสุด และค่าต่ำสุด มิได้พิจารณาการกระจายของข้อมูลที่อยู่ระหว่างคาสูงสุดและค่าต่ำสุดเลย ดังนั้นพิสัยจึงไม่เหมาะสมในการบอกการกระจายของข้อมูลที่มีค่าแตกต่างกัน

2. ความแปรปรวน (variance)

ความแปรปรวนซึ่งแทนด้วยสัญลักษณ์ σ^2 อ่านว่าซิกมากำลังสอง แทนค่าความแปรปรวนที่คิดจากข้อมูลของประชากร เป็นพารามิเตอร์ที่ไม่ทราบค่า ในทางปฏิบัติหาได้เพียงค่าความแปรปรวนที่คิดจากข้อมูลของตัวอย่าง หรือค่าสถิติซึ่งแทนด้วยสัญลักษณ์ s^2 ในเอกสารฉบับนี้จะใช้สัญลักษณ์ s^2

ความแปรปรวนคำนวณจากผลรวมกำลังสองของความแตกต่างระหว่างข้อมูลแต่ละตัวกับค่าเฉลี่ย หารด้วย จำนวนข้อมูลทั้งหมดลบด้วยหนึ่ง อยู่ในรูปสูตรคณิตศาสตร์ ดังนี้

ถ้าให้ $x_1, x_2, x_3, \ldots, x_n$ แทนข้อมูล n ตัว แล้ว

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$$

จากสูตรข้างต้นเขียนในรูปแบบใหม่เพื่อให้ง่ายต่อการคำนวณ จะได้

$$s^2 = \frac{\sum_{i=1}^{n}(x_i)^2 - n(\bar{x})^2}{n-1}$$

เมื่อ i = 1,2,3,...,n และ n แทนจำนวนตัวอย่าง

ถ้าข้อมูลทุกตัวมีค่าเท่ากัน ค่าความแปรปรวนจะเท่ากับศูนย์ ถ้าข้อมูลมีค่าแตกต่างกันมากขึ้น ค่าความแปรปรวนก็จะมากขึ้น แต่เนื่องจากค่าความแปรปรวนของข้อมูลมีหน่วยเป็นหน่วยข้อมูลกำลังสอง จึงทำให้การแปลความหมายของความแปรปรวนนั้นยากต่อการทำความเข้าใจ ดังนั้นเพื่อให้ง่ายต่อการทำความเข้าใจจึงหารากที่สองของความแปรปรวนเพื่อให้หน่วยการกระจายเป็นหน่วยเดียวกับข้อมูล เรียกรากที่สองของความแปรปรวนว่าส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation) เขียนแทนด้วยสัญลักษณ์ s .หรือ sd

$$s = \sqrt{s^2} \qquad \text{หรือ} \qquad s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}} \qquad \text{หรือ} \qquad s = \sqrt{\frac{\sum_{i=1}^{n}(x_i)^2 - n(\bar{x})^2}{n-1}}$$

เมื่อ i = 1,2,3,...,n และ n แทนจำนวนตัวอย่าง

ตัวอย่าง 2.11 จากตัวอย่าง 2.10 จงคำนวณหาค่าความแปรปรวน

วิธีทำ ความแปรปรวนของยอดขายของพนักงานในทีม A

$$s^2 = \frac{\sum_{i=1}^{7}(x_i - \bar{x})^2}{n-1}$$
$$= \frac{(55-55)^2 + (55-55)^2 + \ldots + (55-55)^2}{6}$$
$$s^2 = 0$$

ความแปรปรวนของยอดขายของพนักงานในทีม B

$$s^2 = \frac{\sum_{i=1}^{n}(x_i)^2 - n(\bar{x})^2}{n-1}$$
$$= \frac{(47^2 + 51^2 + 54^2 + \ldots + 63^2) - 7(55)^2}{7-1}$$
$$= \frac{21337 - 21175}{6}$$
$$= \frac{162}{6}$$
$$s^2 = 27$$

ความแปรปรวนของยอดขายของพนักงานในทีม C

$$s^2 = \frac{\sum_{i=1}^{7}(x_i)^2 - n(\bar{x})^2}{n-1}$$
$$= \frac{(39^2 + 47^2 + 53^2 + \ldots + 71^2) - 7(55)^2}{6}$$
$$= \frac{21823 - 21175}{6}$$
$$= \frac{648}{6}$$
$$s^2 = 108$$

ยอดขายของพนักงานทีม A มีค่า $s^2 = 0, s = 0$ หมายความว่ายอดขายของพนักงานแต่ละคนในทีม A ไม่แตกต่างจากยอดขายเฉลี่ย

ทีม B มีค่า $s^2 = 27, s = 5.1962$ หมายความว่ายอดขายของพนักงานแต่ละคนในทีม B มีค่าแตกต่างกันมากกว่าทีม A โดยแตกต่างจากยอดขายเฉลี่ย ± 5.1962 หมื่นบาท

ทีม C มีค่า $s^2 = 108, s = 10.3923$ หมายความว่ายอดขายของพนักงานในทีม C มีค่าแตกต่างกันมากกว่าทีม B และทีม A โดยแตกต่างจากยอดขายเฉลี่ย ± 10.3923 หมื่นบาท

จากการวัดการกระจายของข้อมูลด้วยวิธีการหาค่าความแปรปรวนและส่วนเบี่ยงเบนมาตรฐานนั้น จะเห็นว่าค่าการกระจายของข้อมูลมีความละเอียดมากขึ้น แต่ค่าความแปรปรวนและส่วนเบี่ยงเบนมาตรฐานนั้นจะมีค่าเป็นเลขจำนวนมากหรือน้อยขึ้นอยู่กับค่าของข้อมูลที่สนใจ ถ้าข้อมูลที่สนใจมีค่ามากค่าความแปรปรวนและส่วนเบี่ยงเบนมาตรฐานจะมีค่ามากด้วย ในทางตรงกันข้าม ถ้าข้อมูลที่สนใจมีค่าน้อยค่าความแปรปรวนและส่วนเบี่ยงเบนมาตรฐานจะมีค่าน้อยด้วย นั่นหมายความว่าถ้าต้องการเปรียบเทียบการกระจายของข้อมูลตั้งแต่ 2 กลุ่ม หากเราพิจารณาว่าค่าความแปรปรวนของข้อมูลชุดใดมีค่ามากแสดงว่าข้อมูลกระจายมาก ซึ่งในบางครั้งอาจไม่เป็นเช่นนั้น ค่าความแปรปรวนมากนั้นอาจเป็นเพราะข้อมูลที่นำมาคำนวณมีค่ามากก็ได้ ดังนั้นหากต้องการเปรียบเทียบการกระจายของข้อมูลตั้งแต่ 2 กลุ่ม ควรเปรียบเทียบด้วยค่าสัมประสิทธิ์ความแปรผัน

3. สัมประสิทธิ์ความแปรผัน (coefficient of variation)

สัมประสิทธิ์ความแปรผันซึ่งแทนด้วยสัญลักษณ์ C.V. หมายถึงค่าที่ใช้ในการเปรียบเทียบการกระจายของข้อมูล 2 ชุด หรือมากกว่า ที่มีหน่วยเหมือนกันหรือแตกต่างกัน ซึ่งคำนวณได้จาก

$$C.V. = \frac{s}{\bar{x}} \times 100$$

เมื่อ s แทนส่วนเบี่ยงเบนมาตรฐาน

$\bar{x}$ แทนค่าเฉลี่ย

โดยอาจใช้เกณฑ์พิจารณาเบื้องต้นว่าข้อมูลชุดใด ๆ จะมีระดับความแปรปรวนมาก ค่อนข้างมาก ปานกลาง ค่อนข้างน้อย และน้อย ดังเกณฑ์ต่อไปนี้

สัมประสิทธิ์ความแปรผันของข้อมูล(%)	ระดับความแปรปรวนของข้อมูล
0 – 5	น้อย
6 – 10	ค่อนข้างน้อย
11 – 20	ปานกลาง
21 – 25	ค่อนข้างมาก
ตั้งแต่ 26 ขึ้นไป	มาก

ตัวอย่าง 2.12 ถ้านักเรียนห้องหนึ่งสอบวิชาคณิตศาสตร์ได้คะแนนเฉลี่ย 78 คะแนน ส่วนเบี่ยงเบนมาตรฐาน 8 คะแนน และวิชาภาษาอังกฤษได้คะแนนเฉลี่ย 73 คะแนน ส่วนเบี่ยงเบนมาตรฐาน 7.6 คะแนน อยากทราบว่านักเรียนห้องนี้เรียนวิชาไหนได้คะแนนเกาะกลุ่มกันมากกว่า

จากข้อมูลข้างต้นจะเห็นว่าวิชาคณิตศาสตร์ กับวิชาภาษาอังกฤษนั้นเนื้อหาวิชาไม่เกี่ยวข้องกัน ดังนั้นการพิจารณาการกระจายของคะแนนไม่สามารถพิจารณาจากความแปรปรวนเพียงอย่างเดียวจะต้องพิจารณาสัมประสิทธิ์ความแปรผันด้วย

วิธีทำ จาก $C.V. = \frac{s}{\bar{x}} \times 100$

C.V. วิชาคณิตศาสตร์ $= \frac{8}{78} \times 100 = 10.26\%$

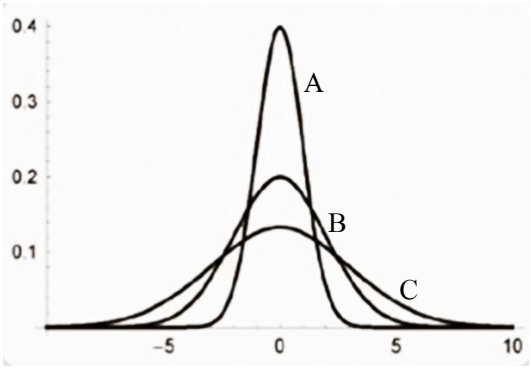
C.V. วิชาภาษาอังกฤษ $= \frac{7.6}{73} \times 100 = 10.41\%$

เนื่องจากค่า C.V. ของคะแนนวิชาคณิตศาสตร์มีค่าน้อยกว่า แสดงว่าการกระจายของคะแนนวิชาคณิตศาสตร์ของนักเรียนในห้องนี้น้อยกว่าการกระจายของคะแนนวิชาภาษาอังกฤษ ดังนั้นนักเรียนในห้องนี้เรียนวิชาคณิตศาสตร์ได้คะแนนเกาะกลุ่มกันมากกว่าวิชาภาษาอังกฤษ

จากวิธีการวัดการกระจายของข้อมูลที่กล่าวมาข้างต้นนั้นเป็นการวัดการกระจายโดยพิจารณาจากตัวเลขที่คำนวณได้ ซึ่งเราสามารถที่จะพิจารณาการกระจายของข้อมูลได้จากลักษณะของโค้งความถี่ หรือการแจกแจงของข้อมูลได้จากความสัมพันธ์ของการกระจายกับการแจกแจงของข้อมูล ดังนี้

4. ความสัมพันธ์ของการกระจายกับการแจกแจงของข้อมูล

พิจารณาโค้งความถี่ของข้อมูล 3 กลุ่ม คือ A B และ C จะเห็นว่าโค้งความถี่ของข้อมูลกลุ่ม A B และ C เป็นโค้งสมมาตร นั้นแสดงว่าข้อมูลทั้ง 3 กลุ่ม มีค่าเป็นไปตามธรรมชาติไม่มีข้อมูลที่ผิดปกติ ค่ากลางของข้อมูลทั้ง 3 กลุ่มจะเท่ากัน แต่เมื่อพิจารณาความสูงของโค้งความถี่จากข้อมูลทั้ง 3 กลุ่ม จะเห็นว่าแตกต่างกัน



ข้อมูลกลุ่มใดที่มีโค้งความถี่สูงแสดงว่าข้อมูลกลุ่มนั้นมีการกระจายน้อย หรือข้อมูลมีค่าแตกต่างกันน้อย และถ้าข้อมูลกลุ่มใดที่มีโค้งความถี่แบนแสดงว่าข้อมูลกลุ่มนั้นมีการกระจายมาก หรือข้อมูลมีค่าแตกต่างกันมาก ดังนั้นข้อมูลกลุ่ม A มีการกระจายน้อยที่สุด ข้อมูลกลุ่ม C มีการกระจายมากที่สุด

ในทางสถิติข้อมูลที่ดี และเหมาะสมที่จะนำไปวิเคราะห์ข้อมูลต่อไปในส่วนของสถิติอ้างอิงนั้นจะต้องมีการกระจายไม่มากหรือน้อยจนเกินไป ซึ่งสามารถพิจารณาการกระจายของข้อมูลจากความโด่งของโค้งความถี่ ดังนี้

5. ความโด่ง (Kurtosis)

ความโด่ง คือความสูงหรือความเตี้ยของโค้งความถี่ของข้อมูลที่ใช้พิจารณาลักษณะการแจกแจงข้อมูลในชุดใด ๆ ความโด่งของโค้งความถี่แบ่งได้เป็น 3 ประเภท ดังนี้

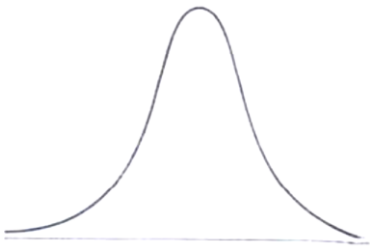
5.1 โค้งความถี่ชนิด Meso Kurtic เป็นโค้งความถี่ที่มีความโด่งเป็นปกติ ซึ่งหมายถึงข้อมูลที่การกระจายเป็นปกติ ลักษณะของโค้งความถี่เป็นดังรูป

Meso Kurtic

5.2 โค้งความถี่ชนิด Platy Kurtic เป็นโค้งความถี่ที่มีความโด่งแบนราบกว่าปกติ ซึ่งหมายถึงข้อมูลที่การกระจายมากกว่าปกติ ลักษณะของโค้งความถี่เป็นดังรูป

Platy Kurtic

5.3 โค้งความถี่ชนิด Lepto Kurtic เป็นโค้งความถี่ที่มีความโด่งสูงกว่าปกติ ซึ่งหมายถึงข้อมูลที่การกระจายน้อยกว่าปกติ ลักษณะของโค้งความถี่เป็นดังรูป



Lepto Kurtic

นอกจากพิจารณาความโด่งจากรูปโค้งความถี่แล้ว เราสมารถวัดความโด่ง (measure of kurtosis) จากค่าสัมประสิทธิ์ความโด่งซึ่งมีวิธีการคำนวณหลายวิธี ในเอกสารฉบับนี้จะใช้ MS EXCEL ในการคำนวณ ความหมายของค่าสัมประสิทธิ์ความโด่งที่คำนวณได้เป็นดังนี้

สัมประสิทธิ์ความโด่ง = 0 แสดงว่าการแจกแจงของข้อมูลมีความโด่งแบบปกติ (Meso Kurtic)

สัมประสิทธิ์ความโด่ง < 0 แสดงว่าการแจกแจงของข้อมูลมีความโด่งราบกว่าปกติ (Platy Kurtic)

สัมประสิทธิ์ความโด่ง > 0 แสดงว่าการแจกแจงของข้อมูลมีความโด่งมากกว่าปกติ (Lepto Kurtic)

ข้อมูลที่ดีและเหมาะสมที่จะใช้ในการวิเคราะห์ข้อมูลต่อไปในส่วนของสถิติอ้างอิงจะต้องไม่มีค่าที่ผิดปกติ และมีความแตกต่างไม่มากหรือน้อยเกินไป หรือมีการแจกแจงแบบปกติ ถ้าพิจารณาด้วยสัมประสิทธิ์ความเบ้และความโด่งควรได้เท่ากับ 0 ซึ่งในทางปฏิบัตินั่นเป็นไปได้ยาก เนื่องจากข้อมูลที่สนใจนั้นมาจากกลุ่มตัวอย่าง ดังนั้นค่าสัมประสิทธิ์ทั้งสองจึงเป็นค่าสถิติ การพิจารณาค่าสัมประสิทธิ์ความเบ้และความโด่งจึงเป็นเพียงการพิจรณาเบื้องต้นว่าข้อมูลมีการแจกแจงแบบปกติหรือไม่เท่านั้น

ตัวอย่าง 2.13 นายตรีรัตน์เป็นนักธุรกิจที่ชอบการเล่นหุ้นมาก ซึ่งในขณะนี้นายตรีรัตน์สนใจหุ้นอยู่ 2 ตัว คือ หุ้น A และหุ้น B นายตรีรัตน์ต้องการตัดสินใจเลือกซื้อหุ้นเพียงตัวเดียวเท่านั้น เพื่อการตัดสินใจอย่างมีเหตุผล นายตรีรัตน์จึงเก็บรวบรวมข้อมูล อัตราผลตอบแทนของหุ้นทั้ง 2 ตัว ในระยะเวลา 5 วัน ได้ผลดังนี้

วันที่	อัตราผลตอบแทนของหุ้น (บาท)	
	หุ้น A	หุ้น B
1	15	14
2	15	15
3	16	15
4	16	17
5	15	16

ถ้าท่านเป็นนายตรีรัตน์จงใช้ข้อมูลที่มีอยู่ประกอบการตัดสินใจในการเลือกซื้อหุ้น โดยมีเงื่อนไขว่าหุ้นนั้นต้องมีผลตอบแทนมากที่สุด และมีความเสี่ยงน้อยที่สุด

วิธีทำ อัตราผลตอบแทนเฉลี่ยของหุ้น A ในระยะเวลา 5 วัน $\bar{x}_A = 15.4$ บาท

อัตราผลตอบแทนเฉลี่ยของหุ้น B ในระยะเวลา 5 วัน $\bar{x}_B = 15.4$ บาท

ความแปรปรวนของอัตราผลตอบแทนของหุ้น A ในระยะเวลา 5 วัน $s_A^2 = 0.3$ บาท^2

ความแปรปรวนของอัตราผลตอบแทนของหุ้น B ในระยะเวลา 5 วัน $s_B^2 = 1.3$ บาท^2

สัมประสิทธิ์ความแปรผันของอัตราผลตอบแทนหุ้น A ในระยะเวลา 5 วัน $C.V._A = 3.57\%$

สัมประสิทธิ์ความแปรผันของอัตราผลตอบแทนหุ้น B ในระยะเวลา 5 วัน $C.V._B = 7.40\%$

สัมประสิทธิ์ความเบ้ของอัตราผลตอบแทนของหุ้น A ในระยะเวลา 5 วัน เท่ากับ 0.61

สัมประสิทธิ์ความเบ้ของอัตราผลตอบแทนของหุ้น B ในระยะเวลา 5 วัน เท่ากับ 0.40

สัมประสิทธิ์ความโด่งของอัตราผลตอบแทนของหุ้น A ในระยะเวลา 5 วัน เท่ากับ -3.33

สัมประสิทธิ์ความโด่งของอัตราผลตอบแทนของหุ้น B ในระยะเวลา 5 วัน เท่ากับ -0.18

จากการวิเคราะห์ข้อมูลข้างต้น จะเห็นว่าอัตราผลตอบแทนเฉลี่ยในระยะเวลา 5 วัน ของหุ้น A และ B เท่ากัน แต่ความแปรปรวน และสัมประสิทธิ์ความแปรผันของอัตราผลตอบแทนของหุ้น B นั้นมากกว่า แสดงว่าอัตราผลตอบแทนของหุ้น B ในเวลา 5 วันนั้นขึ้นและลงมากกว่าหุ้น A หรือความเสี่ยงของหุ้น B มากกว่า หุ้น A นั่นเอง ส่วนสัมประสิทธิ์ความเบ้มีค่าใกล้เคียงกับศูนย์ แสดงว่าอัตรผลตอบแทนหุ้น A และ B ไม่มีค่าผิดปกติ สัมประสิทธิ์ความโด่งติดลบแสดงว่าอัตรผลตอบแทนหุ้น A และ B มีค่าแตกต่างกัน ดังนั้นนายตรีรัตน์ควรเลือกซื้อหุ้น A

ตัวอย่าง 2.14 นายขจรต้องการเลือกซื้อเครื่องชั่งที่มีความเที่ยงตรงในการชั่งน้ำหนักวัว จึงทำการชั่งน้ำหนักวัวตัวหนึ่งด้วยเครื่องชั่ง 2 ประเภท คือเครื่องชั่งแบบดิจิตอล กับเครื่องชั่งแบบตุ้มน้ำหนัก จำนวน 5 ครั้ง ได้ผลดังนี้

ครั้งที่	น้ำหนักวัว (kg)	
	เครื่องชั่งแบบดิจิตอล (A)	เครื่องชั่งแบบตุ้มน้ำหนัก (B)
1	70.1	69.9
2	69.3	70.0
3	69.8	70.0
4	70.1	70.1
5	69.8	70.0

ถ้าท่านเป็นนายขจรท่านจะเลือกซื้อเครื่องชั่งประเภทใด

วิธีทำ ในการชั่งน้ำหนักวัว 5 ครั้งได้ผลการวิเคราะห์สถิติพรรณนา ดังนี้

น้ำหนักเฉลี่ยของวัวที่ชั่งด้วยเครื่องชั่งแบบดิจิตอล $\overline{X}_A$ = 69.82 kg

น้ำหนักเฉลี่ยของวัวที่ชั่งด้วยเครื่องชั่งแบบตุ้มน้ำหนัก $\overline{X}_B$ = 70 kg

ความแปรปรวนของน้ำหนักวัวที่ชั่งด้วยเครื่องชั่งแบบดิจิตอล S_A^2 = 0.107 kg^2

ความแปรปรวนของน้ำหนักวัวที่ชั่งด้วยเครื่องชั่งแบบตุ้มน้ำหนัก S_B^2 = 0.005 kg^2

สัมประสิทธิ์ความแปรผันของน้ำหนักวัวที่ชั่งด้วยเครื่องชั่งแบบดิจิตอล $C.V._A$ = 0.0046

สัมประสิทธิ์ความแปรผันของน้ำหนักวัวที่ชั่งด้วยเครื่องชั่งแบบตุ้มน้ำหนัก $C.V._B$ = 0.0010

สัมประสิทธิ์ความเบ้ของน้ำหนักวัวที่ชั่งด้วยเครื่องชั่งแบบดิจิตอล เท่ากับ -1.15

สัมประสิทธิ์ความเบ้ของน้ำหนักวัวที่ชั่งด้วยเครื่องชั่งแบบตุ้มน้ำหนัก เท่ากับ 0.00

สัมประสิทธิ์ความโด่งของน้ำหนักวัวที่ชั่งด้วยเครื่องชั่งแบบดิจิตอล เท่ากับ 1.33

สัมประสิทธิ์ความโด่งของน้ำหนักวัวที่ชั่งด้วยเครื่องชั่งแบบตุ้มน้ำหนัก เท่ากับ 2.00

จากการวิเคราะห์ข้อมูลข้างต้น จะเห็นว่าในการชั่งน้ำหนักวัวจำนวน 5 ครั้ง เครื่องชั่งแบบตุ้มน้ำหนักชั่งน้ำหนักวัวได้สม่ำเสมอกันมากกว่าเครื่องชั่งแบบดิจิตอล ดังนั้นนายขจรควรเลือกซื้อเครื่องชั่งแบบตุ้มน้ำหนักเพราะมีความเที่ยงตรงมากกว่า

สัดส่วนของสิ่งที่สนใจ (Proportion)

ในบางครั้งสิ่งที่เราสนใจหรือต้องการอธิบายลักษณะข้อมูลอาจไม่ใช่ ค่ากลาง หรือค่าการกระจาย แต่เป็นร้อยละของสิ่งที่สนใจเทียบกับจำนวนทั้งหมด ค่า ๆ นั้นเรียกว่าสัดส่วนของสิ่งที่สนใจ สัญลักษณ์ที่ใช้แทนสัดส่วนของสิ่งที่สนใจในประชากรทั้งหมด คือ p เป็นค่าพารามิเตอร์ที่ไม่ทราบค่า ดังนั้นในทางปฏิบัติจึงหาได้เพียงสัดส่วนของสิ่งที่สนใจในตัวอย่างเป็นค่าสถิติ เขียนแทนด้วยสัญลักษณ์ คือ $\hat{p}$ เราสามารถหาค่าสัดส่วนได้จาก

$$\text{สัดส่วนของสิ่งที่สนใจ}(\hat{p}) = \frac{\text{จานวนสงทสนใจ}}{\text{จำนวนทั้งหมด}}$$

ตัวอย่าง 2.15 จากการสำรวจตัวอย่างเด็กชั้นประถมจำนวน 500 คน ในเขตเมืองหลวง เพื่อตรวจสอบการฉีดวัคซีนป้องกันโรคหัดเยอรมัน พบว่ามีเด็กชั้นประถมจำนวน 150 คนที่ได้รับวัคซีน จากข้อมูลข้างต้นเด็กชั้นประถมที่ได้รับวัคซีนป้องกันโรคหัดเยอรมันคิดเป็นสัดส่วนเท่าใด

วิธีทำ กำหนด $\hat{p}$ แทนสัดส่วนของเด็กชั้นประถมที่ได้รับวัคซีนป้องกันโรคหัดเยอรมัน

$$\hat{p} = \frac{\text{จำนวนเด็กชั้นประถมที่ได้รับวัคซีนป้องกันโรคหัดเยอรมัน}}{\text{จำนวนเด็กชั้นประถมทั้งหมด}}$$

$$= \frac{150}{500}$$

$$= 0.3$$

$\hat{p} = 0.3$ หมายความว่าเด็กชั้นประถมร้อยละ 30 ได้รับวัคซีนป้องกันโรคหัดเยอรมัน

ตัวอย่าง 2.16 บริษัทแห่งหนึ่งผลิตกระถางด้วยวิธีการผลิตที่แตกต่างกัน 2 วิธี จึงต้องการตรวจสอบว่าวิธีการผลิตใดผลิตกระถางได้ดีกว่ากัน จึงเก็บรวบรวมข้อมูลลักษณะของกระถางที่ออกจากเตาพบว่า

วิธีการผลิต	จำนวนกระถางที่ใส่ในเตา(ใบ)	จำนวนกระถางที่แตกในเตาเผา(ใบ)
วิธีที่ 1	200	20
วิธีที่ 2	300	45

จากข้อมูลข้างต้นวิธีการผลิตวิธีใดที่ผลิตกระถางได้ดีกว่ากัน

วิธีทำ กำหนด $\hat{p}_1$ แทนสัดส่วนของกระถางที่แตกในเตา เมื่อผลิตด้วยวิธีที่ 1

$\hat{p}_2$ แทนสัดส่วนของกระถางที่แตกในเตา เมื่อผลิตด้วยวิธีที่ 2

$$\hat{p}_1 = \frac{\text{จำนวนกระถางที่ผลิตด้วยวิธีที่ 1 และแตกในเตา}}{\text{จำนวนกระถางที่ผลิตด้วยวิธีที่ 1}}$$

$$= \frac{20}{200} = 0.10$$

$$\hat{p}_2 = \frac{\text{จำนวนกระถางที่ผลิตด้วยวิธีที่ 2 และแตกในเตา}}{\text{จำนวนกระถางที่ผลิตด้วยวิธีที่ 2}}$$

$$= \frac{45}{300} = 0.15$$

$\hat{p}_1 = 0.10$ หมายความว่ามีกระถางร้อยละ 10 ที่แตกในเตา เมื่อใช้วิธีการผลิตวิธีที่ 1 และ $\hat{p}_2 = 0.15$ หมายความว่ามีกระถางร้อยละ 15 ที่แตกในเตาเมื่อใช้วิธีการผลิตวิธีที่ 2 ดังนั้นควรเลือกวิธีการผลิตวิธีที่ 1

หมายเหตุ สัดส่วนเป็นวิธีที่ใช้ได้กับข้อมูลเชิงปริมาณและเชิงคุณภาพ แต่จากตัวอย่างข้างต้นจะเห็นว่าสัดส่วนมักใช้กับข้อมูลเชิงคุณภาพ หรือข้อมูลในระดับการวัดนามบัญญัติ และเรียงอันดับมากกว่าข้อมูลเชิงปริมาณ

การตรวจสอบข้อมูลก่อนการวิเคราะห์

การวิเคราะห์ข้อมูลในส่วนของสถิติอ้างอิงนั้นต้องใช้ข้อมูลที่เก็บรวบรวมจากตัวอย่างเพื่อหาค่าสถิติ และนำไปอ้างอิง หรือหาข้อสรุปในพารามิเตอร์ที่สนใจ ซึ่งต้องอาศัยขั้นตอน และวิธีการทางสถิติ เพื่อให้การอ้างอิงนั้นมีความถูกต้องแม่นยำมากที่สุดนั้นเอง วิธีการวิเคราะห์ทางสถิติในส่วนของสถิติอ้างอิงนั้น มีด้วยกัน 2 ประเภท ดังนี้

1. สถิติพาราเมตริก (Parametric Statistics)

เป็นการวิเคราะห์ข้อมูลในส่วนของสถิติอ้างอิงที่ถูกคิดค้นขึ้นมาจากการจำลองข้อมูลที่มีการกระจายเป็นแบบปกติ หรือที่เรียกว่าการแจกแจงแบบปกติดังนั้นเทคนิคต่าง ๆ ในสถิติพาราเมตริกจึงมีเงื่อนไขว่าข้อมูลต้องมีการกระจายแบบปกติ

2. สถิตินอนพาราเมตริก (Nonparametric Statistics)

เป็นการวิเคราะห์ข้อมูลในส่วนของสถิติอ้างอิงที่ไม่มีเงื่อนไขว่าข้อมูลต้องมีการกระจายแบบปกติ

ในเอกสารฉบับนี้จะกล่าวถึงวิธีการในส่วนของสถิติพาราเมตริกเท่านั้น ด้วยเหตุนี้ก่อนที่จะนำข้อมูลไปวิเคราะห์ตามวิธีทางสถิติอ้างอิงในเอกสารฉบับนี้ เช่น การประมาณค่า การทดสอบสมมติฐาน ควรตรวจสอบเบื้องต้นก่อนว่าข้อมูลที่เก็บรวบรวมมานั้นมีการกระจายของข้อมูลเป็นแบบปกติหรือไม่ ซึ่งวิธีการตรวจสอบข้อมูลก่อนการวิเคราะห์นี้มีหลายวิธี ดังนี้

1. พิจารณาจากฮิสโตแกรม หรือโค้งความถี่
2. พิจารณาจากค่ากลาง สัมประสิทธิ์ความเบ้ และความโด่ง
3. การทดสอบสมมติฐานเกี่ยวกับการแจกแจงของข้อมูล

แต่ละวิธีมีวิธีการและความเหมาะสมที่แตกต่างกันไป เช่นในบางกรณีหากมีข้อมูลจำนวนไม่มากพอตารางแจกแจงความถี่ ฮิสโตแกรม และโค้งความถี่จะมีรูปร่างไม่ชัดเจน อาจใช้วิธีการหาค่าเฉลี่ย ค่ามัธยฐาน และค่าฐานนิยมของข้อมูลชุดนั้นถ้าทั้ง 3 ค่านั้น มีค่าเท่ากัน หรือใกล้เคียงกัน แสดงว่าข้อมูลที่เก็บรวบรวมมานั้นมีการกระจายเป็นแบบปกติ

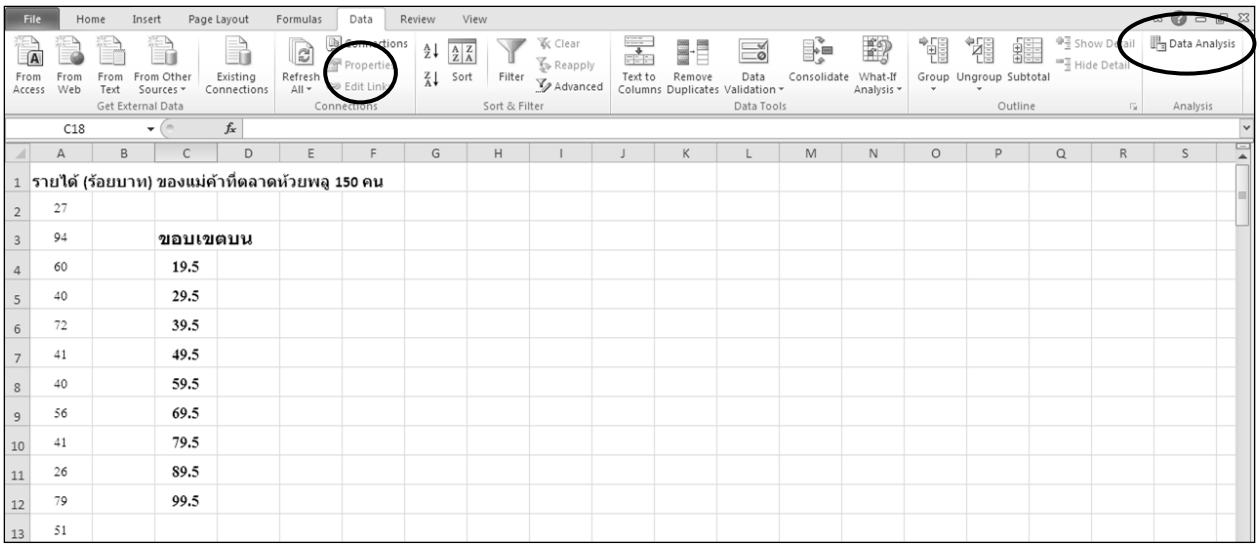
การใช้โปรแกรมสำเร็จรูป

ในส่วนนี้จะใช้ข้อมูลในตัวอย่าง 2.3 เพื่อวิเคราะห์ข้อมูลในส่วนสถิติพรรณนา

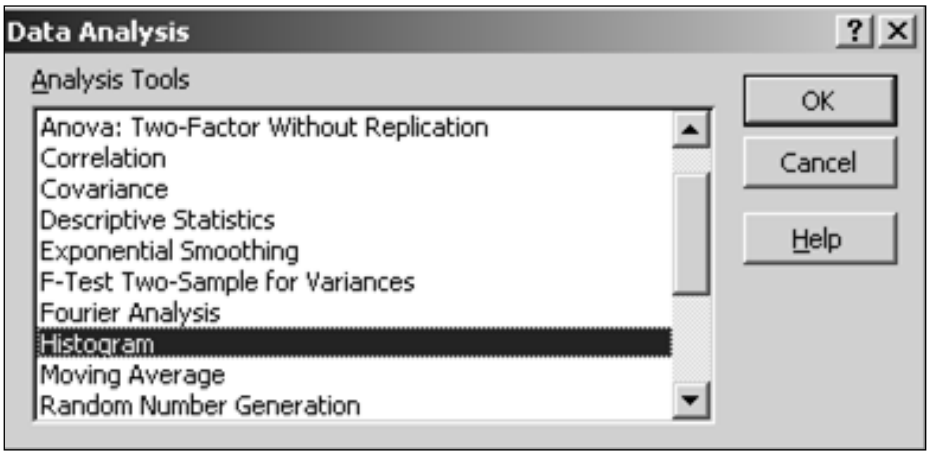
1. การสร้างตารางแจกแจงความถี่ และฮิสโตแกรม

ขั้นตอนที่ 1 ใส่ข้อมูล และขอบเขตบนของตารางแจกแจงความถี่ที่จะสร้าง ไว้ใน cell ที่ต้องการ

เลือกเมนู Data เลือก Data Analysis



ขั้นตอนที่ 2 ในหน้าต่าง Data Analysis เลือก Histogram เลือก OK

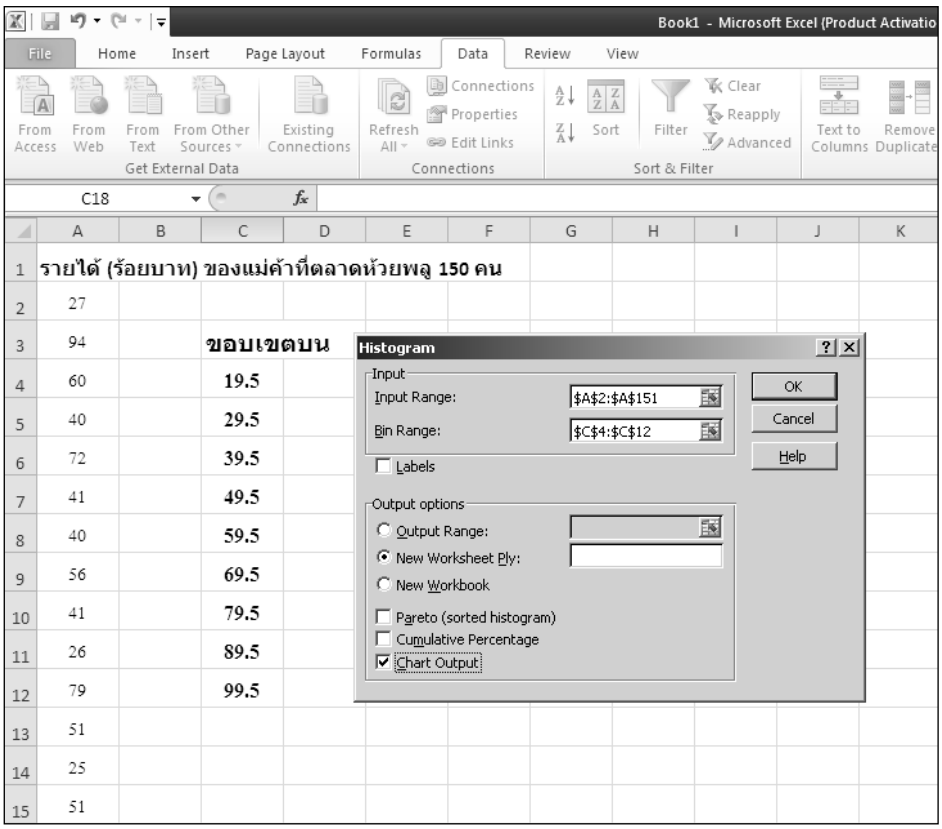


ขั้นตอนที่ 3 ในหน้าต่าง Histogram ในส่วน Input Range ระบุ cell ที่เป็นข้อมูลทั้งหมด

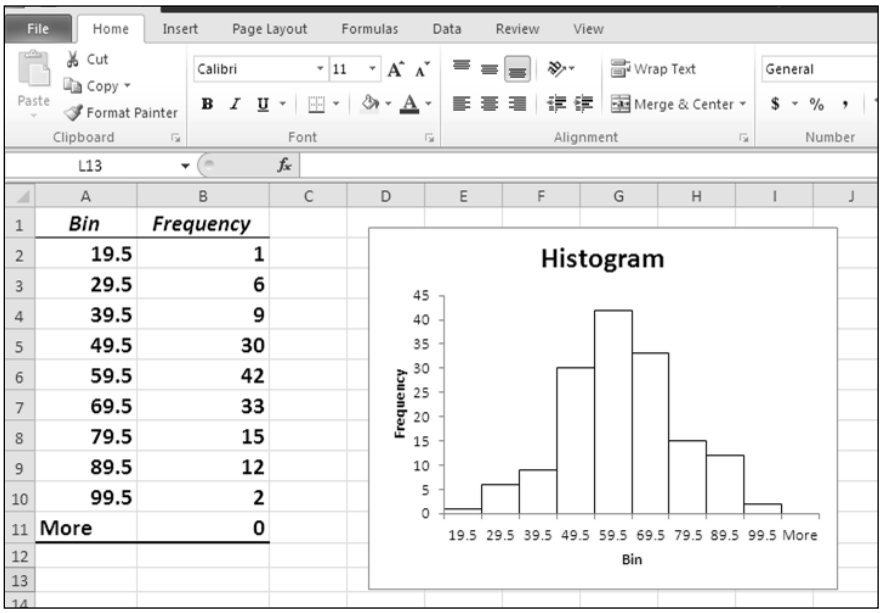
ในส่วน Bin Range ระบุ cell ที่เป็นขอบเขตบน

ในส่วน Output เลือกตำแหน่งผลลัพธ์ และประเภทของแผนภูมิ ดังนี้

- Pareto (sorted histogram) หมายถึงแผนภาพพาเรโต
- Cumulative Percentage หมายถึงความถี่สัมพัทธ์สะสม และกราฟความถี่สะสม
- Chart Output หมายถึงฮิสโตแกรม



ขั้นตอนที่ 4 จะได้ผลลัพธ์ ดังนี้



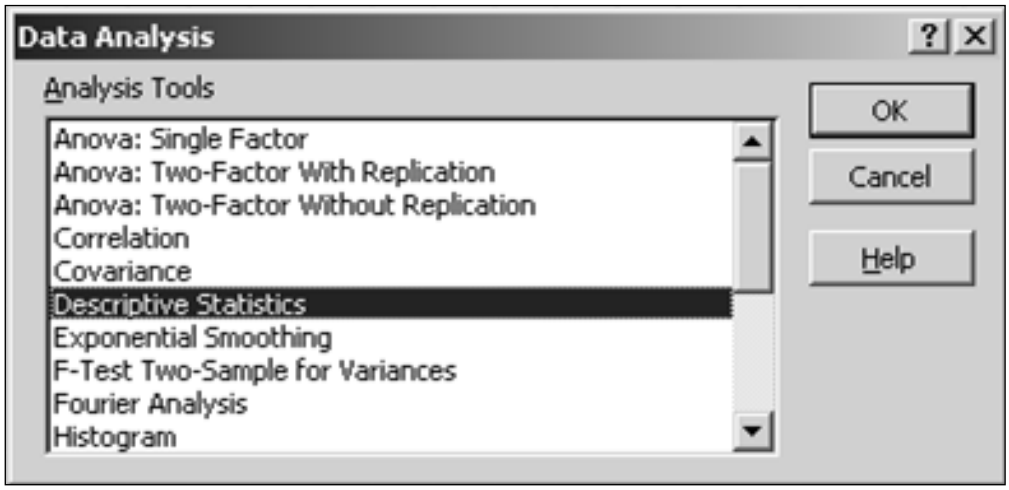
Bin	Frequency
19.5	1
29.5	6
39.5	9
49.5	30
59.5	42
69.5	33
79.5	15
89.5	12
99.5	2
More	0

จากผลลัพธ์ตารางแจกแจงความถี่ คอลัมน์ Bin หมายถึงขอบเขตบนของแต่ละอันตรภาคชั้น เช่นขอบเขตบน 19.5 หมายถึงอันตรภาคชั้น 10-19 ขอบเขตบน 29.5 หมายถึงอันตรภาคชั้น 20-29 เป็นต้น คอลัมน์ Frequency หมายถึงความถี่ของแต่ละอันตรภาคชั้น

จากตารางแจกแจงความถี่พบว่าส่วนใหญ่แม่ค้ามีรายได้ในช่วง 40-69 ร้อยบาท และจากฮิสโตแกรมของรายได้ของแม่ค้าที่ตลาดห้วยพลู 150 คน มีลักษณะคล้ายโค้งสมมาตร แสดงว่ารายได้ของแม่ค้ามีการกระจายเป็นปกติ

2. การวัดแนวโน้มเข้าสู่ส่วนกลาง และการวัดการกระจาย

ขั้นตอนที่ 1 เลือกเมนู Data เลือก Data Analysis เลือก Descriptive Statistics ดังนี้



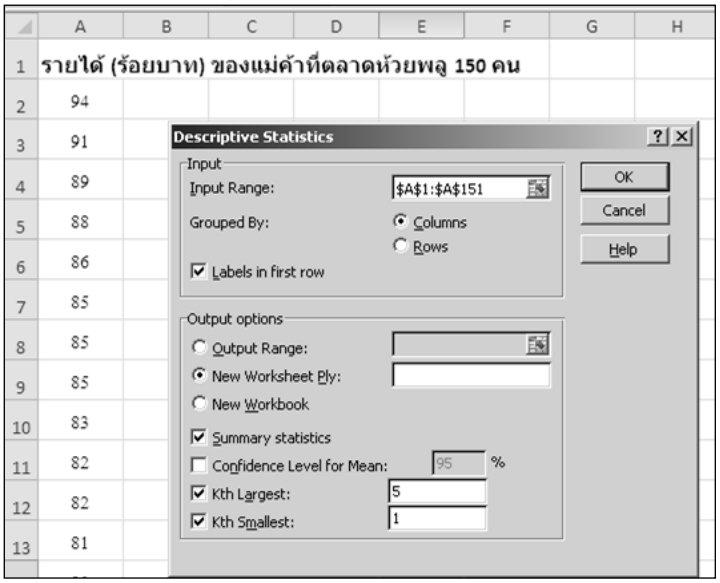
ขั้นตอนที่ 2 ในส่วน Input Range ให้ระบุ cell ที่เป็นข้อมูลทั้งหมด

ในส่วน Group By เป็นการเลือกการแบ่งกลุ่มของข้อมูล

- Columns หมายถึงแบ่งกลุ่มข้อมูลตามคอลัมน์
- Rows หมายถึงแบ่งกลุ่มข้อมูลตามแถว

ในส่วน Output options ระบุตำแหน่งผลลัพธ์ และเลือกการแสดงค่าสถิติ

- Summary statistics หมายถึงแสดงค่าสถิติเบื้องต้น
- Confidence Level for Mean หมายถึงแสดงช่วงความเชื่อมั่นตามที่ระบุ
- Kth Largest หมายถึงแสดงค่ามากสุดตามลำดับที่ระบุ เช่นลำดับ 5
- Kth Smallest หมายถึงแสดงค่าน้อยสุดตามลำดับที่ระบุ เช่นลำดับ 1



ขั้นตอนที่ 3 จะได้ผลลัพธ์ ดังนี้

ค่าเฉลี่ย	56.9066667
ความคลาดเคลื่นมาตรฐาน	1.25076609
มัธยฐาน	56.5
ฐานนิยม	51
ส่วนเบี่ยงเบนมาตรฐาน	15.3186935
ความแปรปรวน	234.662371
สัมประสิทธิ์ความโด่ง	-0.0498428
สัมประสิทธิ์ความเบ้	0.02117849
พิสัย	82
ค่าน้อยสุด	12
ค่ามากสุด	94
ผลรวม	8536
จำนวนข้อมูล	150
ค่ามากสุดลำดับที่ 5	86
ค่าน้อยสุดลำดับที่ 1	12

รายได้(ร้อยบาท) ของแม่ค้าที่ตลาดห้วยพลู 150 คน

หมายความว่าแม่ค้าที่ตลาดห้วยพลู 150 คน มีรายได้เฉลี่ย 56.90 ร้อยบาท แต่ละคนมีรายได้แตกต่างจาก 56.90 ร้อยบาท เท่ากับ 15.31 ร้อยบาท หรือคิดเป็นการกระจาย 26.90%

สรุปท้ายบท

การวิเคราะห์ข้อมูลในส่วนของสถิติพรรณนาเป็นวิธีการวิเคราะห์ข้อมูลเบื้องต้น เพื่อให้ทราบว่าลักษณะของข้อมูลที่เก็บรวบรวมมาในภาพรวม หรือภาพกว้าง ๆ ว่าเป็นอย่างไรตามวัตถุประสงค์ของผู้ใช้ เช่นส่วนใหญ่ข้อมูลที่เก็บรวบรวมมามีค่ามาก หรือน้อย ข้อมูลมีการกระจายหรือมีความแตกต่างกันมากหรือน้อย ข้อมูลที่มีค่าตามที่สนใจมีร้อยละเท่าใดเป็นต้น ซึ่งเทคนิคในการวิเคราะห์ข้อมูลเพื่อตอบวัตถุประสงค์เหล่านี้มีเทคนิคและวิธีการที่แตกต่างกันไป แต่ละเทคนิคแต่ละวิธีจะให้ผลลัพธ์ที่มีความหมายและสารสนเทศที่แตกต่างกัน จึงต้องเลือกเทคนิคและวิธีที่ถูกต้องในการตอบวัตถุประสงค์ และถ้านำผลลัพธ์ในแต่ละเทคนิคแต่ละวิธีมาอธิบายร่วมกันจะทำให้ได้สารสนเทศจากข้อมูลเพื่อใช้ในการประกอบการตัดสินใจ และการวางแผนได้ดียิ่งขึ้น

แบบฝึกหัดท้ายบท

1. จงแจกแจงความถี่ของค่าใช้จ่ายรายวันของนักศึกษา (บาท) ในระดับอุดมศึกษา 80 คน ด้วยแผนภาพลำต้นและใบ ดังนี้

53 57 59 60 60 60 61 61 62 62 62 62 63 63 65 65 65 65 67 67
68 68 68 69 71 71 71 72 72 73 73 73 73 74 74 74 75 75 75 75
75 75 75 76 76 76 76 77 77 78 78 78 78 78 79 79 79 81 81 82
82 83 84 85 85 85 86 87 88 88 88 89 90 93 93 94 95 95 96 97

2. ร้านขายยาแผนปัจจุบันร้านหนึ่งสอบถามอายุ (ปี) ของผู้ที่เข้ามาซื้อยาในร้านในวันหนึ่งจำนวน 40 ราย ปรากฏว่าได้ผลดังนี้ 53 58 61 64 40 34 21 12 40 37 24 13 42 38 28 13 43 38 31 16 10 21 33 39 52 52 39 33 21 10 49 7 19 20 31 32 39 39 47 48 จงสร้าง

   2.1 ตารางแจกแจงความถี่และความถี่สัมพัทธ์ โดยให้มีอันตรภาคชั้น 6 อันตรภาคชั้น

   2.2 จากตารางข้อ 2.1. จงสร้างฮิสโตแกรม รูปหลายเหลี่ยมความถี่ และโค้งความถี่

3. จงหาค่าเฉลี่ย มัธยฐาน และฐานนิยม ของข้อมูลต่อไปนี้

3.1 7 4 10 9 15 12 7 9 7

3.2 8 11 4 3 2 4 10 6 4 1 10 7 12 6 6 10

3.3 129 114 169 141 167 144 123 105 134 132

4. ข้อมูลต่อไปนี้ถูกวัดด้วยระดับใด ค่ากลางที่เหมาะสมที่สุดคือค่ากลางวิธีใด และมีค่าเท่าไร

4.1 ร้านค้าแห่งหนึ่งต้องการทราบเกี่ยวกับยอดเงินค้างชำระเฉลี่ยของลูกค้า 15 คน ซึ่งมียอดค้างชำระ (บาท) ของแต่ละราย ดังนี้ 55.20 4.88 271.95 18.06 180.29 365.29 28.16 399.11 807.80 44.14 97.47 9.98 61.61 56.89 82.73

4.2 บริษัทแห่งหนึ่งต้องการสำรวจว่าพนักงานในบริษัทส่วนใหญ่มีวุฒิการศึกษาอยู่ในระดับใด จึงทำการเก็บข้อมูลโดยกำหนดให้ 1 หมายถึงการศึกษาระดับประถมหรือต่ำกว่า 2 หมายถึงการศึกษาระดับมัธยมศึกษา 3 หมายถึงการศึกษาระดับปริญญาตรี 4 หมายถึงการศึกษาระดับปริญญาโท 5 หมายถึงการศึกษาระดับปริญญาเอก ปรากฏผลดังนี้ 1 5 2 4 3 2 5 4 2 1 3 1 2 3 3 4 5 3 3 3

4.3 ขส.มก. ต้องการทราบว่ารถประจำทางสายใดที่มีผู้ใช้บริการมากที่สุด จึงทำการเก็บข้อมูลปรากฏผลดังนี้

สายรถประจำทาง	1	2	3	4	5	6
จำนวนรถ	21	38	18	30	38	19

5. บริษัทโตโยต้านครปฐม ต้องการให้เงินสมนาคุณแก่พนักงานขายคนใดคนหนึ่งในจำนวน 3 คน ที่มียอดขายรวม 5 ปีอยู่ในระดับแนวหน้า ดังข้อมูลต่อไปนี้ (หน่วย: 100,000 บาท)

ประณต	74	80	72	65	78
ประดิษฐ์	82	76	69	70	84
ประเสริฐ	77	80	75	69	73

ถ้าบริษัทจะให้เงินสมนาคุณดังกล่าวแก่พนักงานที่มียอดขายคงเส้นคงวามากที่สุด จงใช้เหตุผลจากการวิเคราะห์ข้อมูลเบื้องต้นเพื่อสรุปว่าพนักงานขายคนใดจะได้รับเงินนั้น

6. จากการสำรวจธนบัตรที่ชำรุดจากการใช้งาน จำนวน 15100 ฉบับ พบว่าเกิดมาจากสาเหตุต่าง ๆ คือขาด 2580 ใบ มีรอยที่เย็บกระดาษ 5120 ใบมีรอยขีดเขียน 400 ใบ และมีรอยพับ 7000 ใบ ทางธนาคารแห่งประเทศไทยสนใจว่าธนบัตรที่ชำรุดเนื่องจากมีรอยที่เย็บกระดาษ และมีรอยพับคิดเป็นสัดส่วนเท่าไร