บทที่ 4

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

การแจกแจงความถี่ (frequency distribution)

เทคนิคการแจกแจงความถี่ เป็นเทคนิคที่จำแนกข้อมูลตามลักษณะต่าง ๆ ที่ซ้ำ ๆ แล้วนับจำนวน หรือเรียกว่าความถี่ (frequency) จึงทำให้ทราบว่าข้อมูลส่วนใหญ่มีค่าเป็นเท่าไรและนอกจากนี้ยังเป็นการจัดเตรียมข้อมูลเพื่อความสะดวกในการวิเคราะห์ข้อมูลขั้นต่อไป ส่วนวิธีการในการแจกแจงความถี่มีหลายวิธี ในเอกสารฉบับนี้จะกล่าวถึงการแจกแจงความถี่ด้วยตาราง หรือเรียกว่าตารางแจกแจงความถี่ ซึ่งประกอบด้วยตารางแจกแจงความถี่จำแนกทางเดียว กับตารางแจกแจงความถี่จำแนกสองทาง

1. ตารางแจกแจงความถี่จำแนกทางเดียว (One Way Frequency Table)

ตารางแจกแจงความถี่จำแนกทางเดียวเป็นตารางแจกแจงความถี่ที่แยกหรือจำแนกข้อมูลที่มีค่าซ้ำ ๆ กัน ตามลักษณะของตัวแปรเดียว

ตัวอย่าง 4.1 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
9	ญ	11

คนที่	เพศ	น้ำหนักกระเป๋า (kg)
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 4.1 จำนวนผู้โดยสารเครื่องบินเที่ยวกรุงเทพ ฯ - เชียงใหม่จำแนกตามเพศ

เพศ	ความถี่
ชาย	11
หญิง	9
	20

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

เนื่องจากน้ำหนักสัมภาระเป็นข้อมูลเชิงปริมาณมีค่าต่ำสุดเป็น 7 และค่าสูงสุดเป็น 11 ซึ่งแตกต่างกันไม่มากนัก และมีข้อมูลบางค่าที่ซ้ำกันจึงสร้างตารางแจกแจงความถี่ได้ ดังนี้

ตารางที่ 4.2 จำนวนผู้โดยสารเครื่องบินเที่ยวกรุงเทพ ฯ – เชียงใหม่จำแนกตามน้ำหนักสัมภาระ (kg)

น้ำหนักสัมภาระ(kg)	ความถี่
7	5
8	5
9	2
10	4
11	4
	20

จากตารางแจกแจงความถี่หมายความว่า จากผู้โดยสารเครื่องบินเที่ยวกรุงเทพ ฯ ถึงเชียงใหม่จำนวน 20 คน ส่วนใหญ่นำสัมภาระติดตัวมาหนัก 7 - 8 กิโลกรัม

หมายเหตุ ในกรณีที่ข้อมูลเชิงปริมาณมีค่าแตกต่างกันมาก คอลัมภ์แรกของตารางแจกแจงความถี่ชนิดไม่จัดข้อมูลเป็นหมวดหมู่ที่แสดงแต่ละค่าของข้อมูลก็จะมีความยาวมากเกินไป ดังนั้นตารางแจกชนิดนี้จึงเหมาะสมกับข้อมูลเชิงคุณภาพมากกว่า แต่เราก็สามารถแจกแจงความถี่ของข้อมูลเชิงปริมาณในรูปตารางแจกแจงความถี่ชนิดจัดข้อมูลเป็นหมวดหมู่ได้ โดยแบ่งข้อมูลเชิงปริมาณเป็นช่วง ๆ เสียก่อนด้วยการเปลี่ยนรหัส (recode) ดังตาราง 4.3

ตารางที่ 4.3 จำนวนผู้โดยสารเครื่องบินเที่ยวกรุงเทพ ฯ - เชียงใหม่จำแนกตามน้ำหนักสัมภาระ (kg)

น้ำหนักสัมภาระ (kg)	ความถี่
7-9	2
10-12	8
13-15	14
16-18	19
19-21	7
	50

จากตารางแจกแจงความถี่หมายความว่า จากผู้โดยสารเครื่องบินเที่ยวกรุงเทพ ฯ ถึงเชียงใหม่จำนวน 50 คน ส่วนใหญ่นำสัมภาระติดตัวมาหนัก 13 - 18 กิโลกรัม

2. ตารางแจกแจงความถี่จำแนกสองทาง (Two Way Frequency Table)

เป็นตารางแจกแจงความถี่ที่แยกหรือจำแนกข้อมูลที่มีค่าซ้ำ ๆ กัน ตามลักษณะของตัวแปร 2 ตัว

ตัวอย่าง 4.2 จากข้อมูลในตัวอย่าง 4.1 จงสร้างตารางแจกแจงความถี่ของผู้โดยสารเครื่องบินเที่ยวกรุงเทพฯ - เชียงใหม่จำนวน 20 คน จำแนกตามเพศ และน้ำหนักสัมภาระ

ตารางที่ 4.3 จำนวนผู้โดยสารเครื่องบินเที่ยวกรุงเทพ ฯ - เชียงใหม่จำแนกตามเพศและน้ำหนักสัมภาระ (kg)

เพศ	น้ำหนักสัมภาระ (kg)					รวม
	7	8	9	10	11	
ชาย	5	3	1	0	2	11
หญิง	0	2	1	4	2	9
รวม	5	5	2	4	4	20

จากตารางแจกแจงความถี่หมายความว่า จากผู้โดยสารเครื่องบินเที่ยวกรุงเทพ ฯ ถึงเชียงใหม่จำนวน 20 คน ส่วนใหญ่ที่นำสัมภาระติดตัวมาหนัก 10 - 11 กิโลกรัมจะเป็นเพศหญิง และนำสัมภาระติดตัวมาหนัก 7 – 8 กิโลกรัมจะเป็นเพศชาย

3 การแจกแจงความถี่ด้วยกราฟ (graphing frequency distribution)

การแจกแจงความถี่ด้วยกราฟเป็นการแจกแจงความถี่ต่อจากตารางแจกแจงความถี่ในลักษณะที่เป็นรูปกราฟ ซึ่งทำได้ 3 วิธี คือ

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

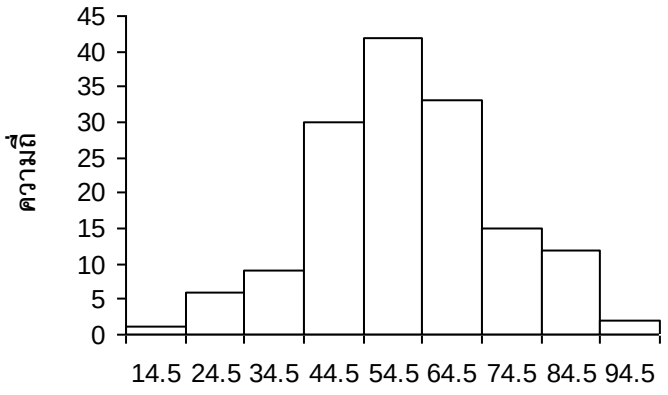
3.1 ฮิสโตแกรม การนำเสนอข้อมูลวิธีนี้มีลักษณะเป็นการแสดงความถี่ของข้อมูลด้วยพื้นที่ของรูปสี่เหลี่ยมผืนผ้าที่สร้างขึ้นติดต่อกันบนแกนนอน ดังนั้นผลรวมของพื้นที่ของรูปสี่เหลี่ยมผืนผ้าทุกแท่งจะเท่ากับจำนวนข้อมูลทั้งหมดที่นำมาแจกแจง

ตัวอย่าง 4.3 จากตารางแจกแจงความถี่แสดงจำนวนแม่ค้าจำแนกตามรายได้ 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

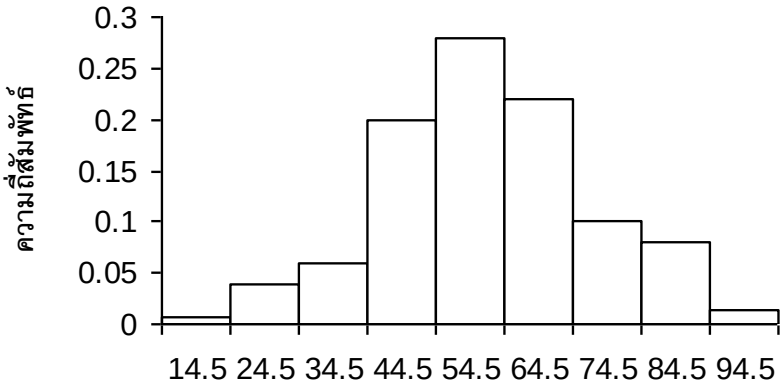
อิสโตแกรมความถี่ของรายได้



จากรูปฮิสโตแกรมที่แกน X แสดงค่าข้อมูล (หรืออันตรภาคชั้น) และแกน Y แสดงความถี่เรียกว่าฮิสโตแกรมความถี่ ซึ่งมีพื้นที่ทุกแท่งรวมกันเท่ากับความถี่รวม

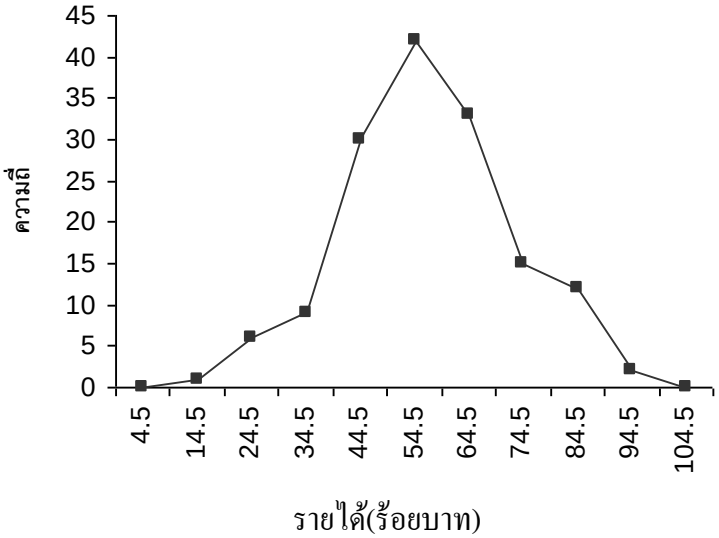
ส่วนฮิสโตแกรมที่แกน X แสดงค่าข้อมูล (หรืออันตรภาคชั้น) และแกน Y แสดงความถี่สัมพัทธ์ หรือร้อยละเรียกว่าฮิสโตแกรมความถี่สัมพัทธ์ (relative frequency histogram) หรือฮิสโตแกรมร้อยละ (percentage histogram) ซึ่งมีพื้นที่ทุกแท่งรวมกันเท่ากับ 1 หรือ 100% ดังนี้

อิสโตแกรมความถี่สัมพัทธ์ของรายได้



3.2 รูปหลายเหลี่ยมความถี่ คือกราฟเส้นที่ได้จากการลากเส้นตรงเชื่อมจุดกึ่งกลางปลายยอดทุกแท่งของฮิสโตแกรม ซึ่งเป็นผลมาจากรูปฮิสโตแกรม ดังนี้

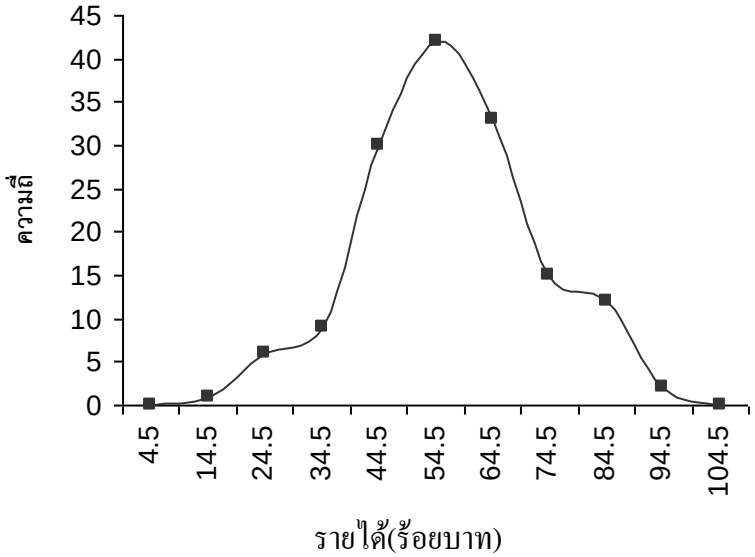
รูปหลายเหลี่ยมความถี่ของรายได้



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

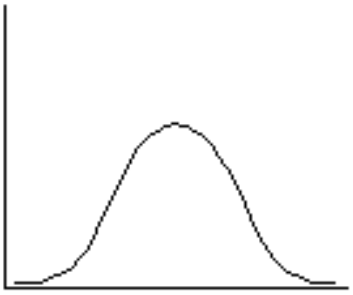
รูปโค้งความถี่ของรายได้



4. ชนิดของโค้งความถี่

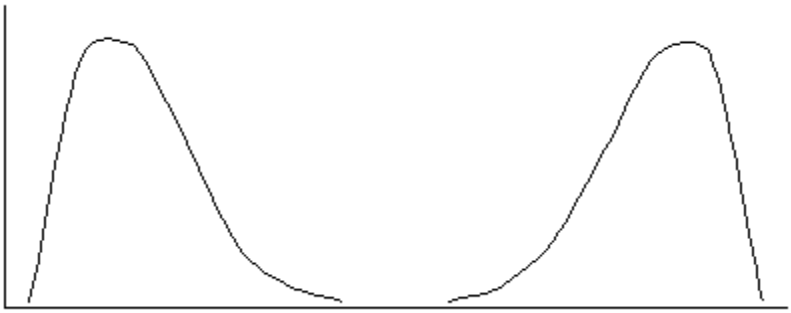
จากโค้งความถี่ หรือโค้งความถี่สัมพัทธ์ บอกให้ทราบถึงลักษณะการแจกแจงความถี่ของข้อมูล สามารถแบ่งตามรูปร่างของโค้งความถี่ได้ 6 ชนิด คือ

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 เป็นค่าพารามิเตอร์ที่ไม่ทราบค่า 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 4.4 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

$$\bar{x} = \frac{\sum_{i=1}^{4} x_i}{4}$$

$$= \frac{11.30 + 12.10 + 13.45 + 12.40}{4}$$

$$= 12.31$$

ราคาน้ำมันเบนซินเฉลี่ยต่อลิตรใน 4 เดือนราคา 12.31 บาท

หมายเหตุ 1. ค่าเฉลี่ยนี้ใช้กับข้อมูลเชิงปริมาณ

2. ค่าเฉลี่ยเหมาะสมกับข้อมูลที่มีค่าสอดคล้องกัน นั่นคือไม่มีค่าผิดปกติ (ค่าที่มากหรือน้อยจนเกินไป) หรือมีโค้งความถี่เป็นโค้งสมมาตร

3. ค่าเฉลี่ยเหมาะสมกับข้อมูลที่มีจำนวนมาก ถ้าตัดค่าที่มากหรือน้อยเกินไปแล้วจะไม่มีผลต่อค่าเฉลี่ย

2 มัธยฐาน (median)

มัธยฐานซึ่งแทนด้วยสัญลักษณ์ Me หมายถึงค่ากลางที่มีตำแหน่งอยู่ตรงกลางของข้อมูลที่จัดเรียงจากน้อยไปหามาก หรือมากไปหาน้อยเรียบร้อยแล้ว ค่ามัธยฐานจึงแบ่งข้อมูลออกเป็น 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 4.5 จากการสำรวจคุณภาพของสินค้าชนิดหนึ่งจากผู้บริโภค 10 คน กำหนดให้คะแนนเต็ม 5 คะแนน โดย 0 หมายถึงคุณภาพไม่ดี 5 หมายถึงคุณภาพดีมาก ตามลำดับ ได้ผลดังนี้ 1 4 3 5 1 2 2 4 5 4 จงหาค่ากลางของข้อมูลชุดนี้โดยใช้ค่ามัธยฐาน

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 $\frac{3+4}{2} = \frac{7}{2} = 3.5$ ค่ามัธยฐาน = 3.5 หมายความว่าคุณภาพของสินค้าชนิดนี้อยู่ในระดับคะแนน 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 4.6 จากการสำรวจของบริษัทแห่งหนึ่งเกี่ยวกับการเลือกอาชีพหลังจบการศึกษาปริญญาตรีของนักศึกษา 17 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 จงหาฐานนิยมของข้อมูลชุดนี้

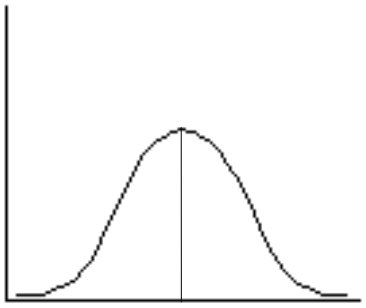
อาชีพ A มีความถี่ 1
อาชีพ B มีความถี่ 2
อาชีพ C มีความถี่ 2
อาชีพ D มีความถี่ 3
อาชีพ E มีความถี่ 2
อาชีพ F มีความถี่ 5
อาชีพ G มีความถี่ 2
อาชีพ F มีความถี่มากที่สุด ดังนั้นค่ากลางฐานนิยมของข้อมูลชุดนี้ อาชีพ F

หมายเหตุ ฐานนิยมสามารถใช้ได้กับข้อมูลเชิงคุณภาพ และเชิงปริมาณ แต่เหมาะสมกับข้อมูลเชิงคุณภาพมากกว่า เพราะเป็นการละทิ้งข้อสนเทศบางอย่างที่มีอยู่ในข้อมูลไป

การหาค่ากลางของข้อมูลชุดเดียวกันโดยวิธีต่าง ๆ ไม่จำเป็นต้องจะได้ค่ากลางแต่ละค่าที่มีค่าเท่ากัน ขึ้นอยู่กับลักษณะของข้อมูลที่นำมาวิเคราะห์

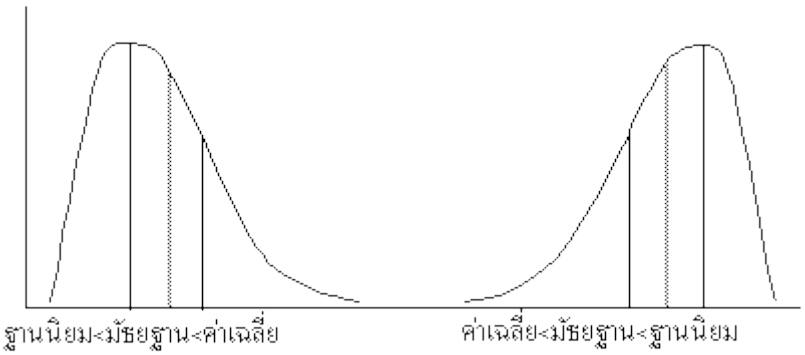
4 ความสัมพันธ์ระหว่างค่าเฉลี่ย, มัธยฐาน, ฐานนิยม

ถ้าลักษณะของข้อมูลที่นำมาวิเคราะห์มีการแจกแจงสมมาตร หรือโค้งความถี่เป็นรูปสมมาตร (โค้งปกติ) ค่าเฉลี่ย ค่ามัธยฐาน และค่าฐานนิยมจะมีค่าเท่ากัน ดังนี้



ค่าเฉลี่ย
มัธยฐาน
ฐานนิยม

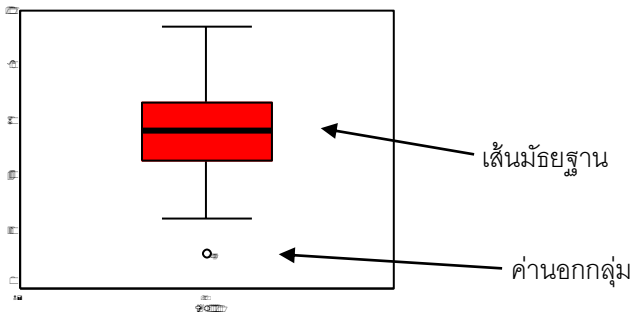
ถ้าลักษณะของข้อมูลที่นำมาวิเคราะห์มีการแจกแจงแบบเบ้ขวา หรือเบ้ซ้ายค่าเฉลี่ย ค่ามัธยฐาน และค่าฐานนิยมจะมีค่า ดังนี้



จากรูปโค้งเบ้ขวา แสดงว่าข้อมูลส่วนมากมีค่าน้อย ดังนั้นฐานนิยมจะอยู่ ณ ตำแหน่งที่โค้งสูงที่สุด และเนื่องจากโค้งเบ้ขวาแสดงว่ามีข้อมูลบางค่าที่มีค่ามากเกินไปจึงมีผลกระทบทำให้ค่าเฉลี่ยมีค่ามากกว่าค่าฐานนิยม แต่ไม่มีผลกระทบต่อค่ามัธยฐานจึงทำให้มัธยฐานมีค่าน้อยกว่าค่าเฉลี่ย แต่มากกว่าฐานนิยม และเนื่องจากมัธยฐานเป็นค่าที่แบ่งพื้นที่ใต้โค้งออกเป็น 2 ส่วนเท่า ๆ กัน ดังนั้น กรณีโค้งเบ้ขวาจะได้ว่า $Mo < Me < \bar{x}$ และในทำนองเดียวกันกรณีโค้งเบ้ซ้ายจะได้ว่า

$\bar{x} < Me < Mo$

บล็อกพลอต (Boxplot) เป็นกราฟที่แสดงรายละเอียดของการกระจายของข้อมูล สามารถระบุได้ว่าข้อมูลเบ้ซ้าย เบ้ขวา หรือสมมาตร ข้อมูลนั้นมีค่ามาก หรือน้อยผิดปกติหรือไม่ และยังสามารถค่ากลางและค่าการกระจายของข้อมูลหลาย ๆ ชุดได้



จากข้อมูลคะแนนสอบวิชาสถิติของนักศึกษา 150 คน จากบล็อกพลอตพบว่ามีตัวอย่างหน่วยหนึ่งที่ได้คะแนนสอบน้อยมากเมื่อเทียบกับคนอื่น ๆ เส้นมัธยฐานอยู่ค่อนข้างกลางกล่อง ประมาณได้ว่าคะแนนส่วนใหญ่อยู่ใกล้ ๆ ค่ากลาง มีนักศึกษาได้คะแนนน้อย หรือมากกว่าค่ากลางในจำนวนเท่า ๆ กัน

5. ความเบ้ (Skewness)

ความเบ้ คือระดับความเอนเอียงหรือความไม่สมมาตรของการแจกแจงของข้อมูล โดยพิจารณาได้จากโค้งความถี่ หรือวัดความเบ้ (measure of skewness) จากค่าสัมประสิทธิ์ความเบ้ ซึ่งมีวิธีการคำนวณหลายวิธี 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 ถ้าข้อมูลแตกต่างกันมาถือว่ามีการกระจายมาก ถ้าข้อมูลแตกต่างกันน้อยถือว่ามีการกระจายน้อย ดังตัวอย่าง 4.7

ตัวอย่าง 4.7 ในการคัดเลือกพนักงาน 1 ทีมเพื่อรับรางวัลทีมงานที่มีประสิทธิภาพในการขายดีเด่น โดยพิจารณาจากยอดขาย (หมื่นบาท) จากพนักงานแต่ละทีม ดังนี้

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

จากยอดขายของพนักงานทั้ง 3 ทีมนี้ ถ้าพิจารณาที่ค่าเฉลี่ยจะพบว่าพนักงานทั้ง 3 ทีมมียอดขายเฉลี่ยเท่ากับ 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 4.7 พิสัยของข้อมูลชุดที่ 1 = 55 - 55 = 0

พิสัยของข้อมูลชุดที่ 2 = 63 - 47 = 16

พิสัยของข้อมูลชุดที่ 3 = 71 - 39 = 32

จากค่าพิสัยจะเห็นว่าข้อมูลชุดที่ 3 มีการกระจายมากที่สุด ตามด้วยชุดที่ 2 และชุดที่ 1

จะเห็นว่าการวัดการกระจายโดยใช้พิสัยนั้นเป็นวิธีที่ง่ายและรวดเร็ว แต่เป็นค่าวัดการกระจายแบบหยาบ ๆ เนื่องพิจารณาจากค่าเพียง 2 ค่าเท่านั้น คือค่าสูงสุด และค่าต่ำสุด แต่มิได้พิจารณาการกระจายของข้อมูลที่อยู่ระหว่างคาสูงสุดและค่าต่ำสุดเลย ดังนั้นพิสัยจึงเป็นค่าที่อาจไม่เหมาะสมนักในการบอกการกระจายของข้อมูล

2 ความแปรปรวน (variance)

ความแปรปรวนซึ่งแทนด้วยสัญลักษณ์ σ^2 อ่านว่าซิกมากำลังสอง แทนค่าความแปรปรวนที่คิดจากข้อมูลของประชากร เป็นค่าพารามิเตอร์ที่ไม่ทราบค่า ดังนั้นในทางปฏิบัติจึงหาได้เพียงค่าความแปรปรวนที่คิดจากข้อมูลของตัวอย่าง หรือค่าสถิติซึ่งแทนด้วยสัญลักษณ์ s^2 ในเอกสารฉบับนี้จะใช้สัญลักษณ์ s^2

ความแปรปรวนคำนวณจากผลรวมกำลังสองของความแตกต่างระหว่างข้อมูลแต่ละตัวกับค่าเฉลี่ย หารด้วย จำนวนข้อมูลทั้งหมดลบด้วยหนึ่ง อยู่ในรูปสูตรคณิตศาสตร์ ดังนี้

ถ้าให้ $x_1, x_2, x_3, \ldots, x_k$ 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 n แทนจำนวนข้อมูล

ถ้าข้อมูลทุกตัวมีค่าเท่ากัน ค่าความแปรปรวนจะเท่ากับศูนย์ ถ้าข้อมูลมีค่าแตกต่างกันมากขึ้น ค่าความแปรปรวนก็จะมากขึ้น

รากที่สองของความแปรปรวนเรียกว่าส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation) เขียนแทนด้วยสัญลักษณ์ s .

$$s = \sqrt{s^2}$$

ตัวอย่าง 4.8 จากตัวอย่าง 4.7 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

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

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

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

ทีม C มีค่า $s^2 = 108, s = 10.3923$ หมายความว่ายอดขายของพนักงานในทีม C มีการกระจายของมากกว่าทีม B และทีม A โดยแตกต่างจากยอดขายเฉลี่ย $\pm$ 10.3923 หมื่นบาท

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 4.9 ถ้านักเรียนห้องหนึ่งสอบวิชาคณิตศาสตร์ได้คะแนนเฉลี่ย 78 คะแนน ส่วนเบี่ยงเบนมาตรฐาน 8 คะแนน และวิชาภาษาอังกฤษได้คะแนนเฉลี่ย 73 คะแนน ส่วนเบี่ยงเบนมาตรฐาน 7.6 คะแนน อยากทราบว่านักเรียนห้องนี้เรียนวิชาไหนได้คะแนนเกาะกลุ่มกันมากกว่า

จากข้อมูลข้างต้นจะเห็นว่าวิชาคณิตศาสตร์ กับวิชาภาษาอังกฤษนั้นเนื้อหาวิชาไม่เกี่ยวข้องกัน ดังนั้นการพิจารณาการกระจายของคะแนนไม่สามารถพิจารณาจากความแปรปรวนเพียงอย่างเดียวจะต้องพิจารณาสัมประสิทธิ์ความแปรผันด้วย

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

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

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

เนื่องจาก C.V. ของคะแนนวิชาคณิตศาสตร์มีค่าน้อยกว่า แสดงว่าการกระจายของคะแนนวิชาคณิตศาสตร์ของนักเรียนในห้องนี้น้อยกว่าคะแนนวิชาภาษาอังกฤษ ดังนั้นนักเรียนในห้องนี้เรียนวิชาคณิตศาสตร์ได้คะแนนเกาะกลุ่มกันมากกว่าวิชาภาษาอังกฤษ

ตัวอย่าง 4.10 นายขจรต้องการเลือกซื้อเครื่องชั่งน้ำหนักที่มีความเที่ยงตรงในการชั่งน้ำหนักนักเรียน จึงทำการชั่งน้ำหนักนักเรียนแต่ละคนหนึ่งด้วยเครื่องชั่ง 2 ประเภท คือเครื่องชั่งแบบดิจิตอล กับเครื่องชั่งแบบตุ้มน้ำหนัก จำนวน 5 คน ได้ผลดังนี้

คนที่	น้ำหนักนักเรียน (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

ในการชั่งน้ำหนักนักเรียน 5 คนได้ผลการวิเคราะห์สถิติพรรณนา ดังนี้
น้ำหนักเฉลี่ยของนักเรียนที่ชั่งด้วยเครื่องชั่งแบบดิจิตอล $\overline{X}_A$ = 69.82 kg
น้ำหนักเฉลี่ยของนักเรียนที่ชั่งด้วยเครื่องชั่งแบบตุ้มน้ำหนัก $\overline{X}_B$ = 70 kg
ความแปรปรวนของน้ำหนักนักเรียนที่ชั่งด้วยเครื่องชั่งแบบดิจิตอล S_A^2 = 0.107 kg^2
ความแปรปรวนของน้ำหนักนักเรียนที่ชั่งด้วยเครื่องชั่งแบบตุ้มน้ำหนัก S_B^2 = 0.005 kg^2
สัมประสิทธิ์ของความแปรผันของน้ำหนักนักเรียนที่ชั่งด้วยเครื่องชั่งแบบดิจิตอล $C.V._A$ = 0.0046
สัมประสิทธิ์ของความแปรผันของน้ำหนักนักเรียนที่ชั่งด้วยเครื่องชั่งแบบตุ้มน้ำหนัก $C.V._B$ = 0.0010

จากการวิเคราะห์ข้อมูลข้างต้น จะเห็นว่าในการชั่งน้ำหนักนักเรียนจำนวน 5 คน เครื่องชั่งแบบตุ้มน้ำหนักชั่งน้ำหนักนักเรียนได้สม่ำเสมอกันมากกว่าเครื่องชั่งแบบดิจิตอล ดังนั้น นายขจรควรเลือกซื้อเครื่องชั่งแบบตุ้มน้ำหนักเพราะมีความเที่ยงตรงมากกว่า

การใช้โปรแกรมสำเร็จรูป SPSS

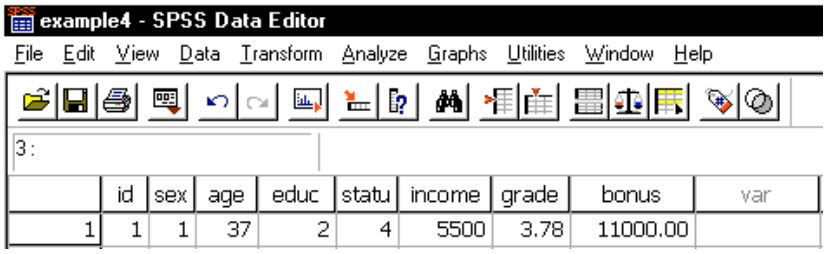
ในส่วนนี้จะใช้ข้อมูลในแฟ้มข้อมูล exsample4.sav เพื่อวิเคราะห์ข้อมูลในส่วนสถิติพรรณนา

1. การสร้างตารางแจกแจงความถี่จำแนกทางเดียว

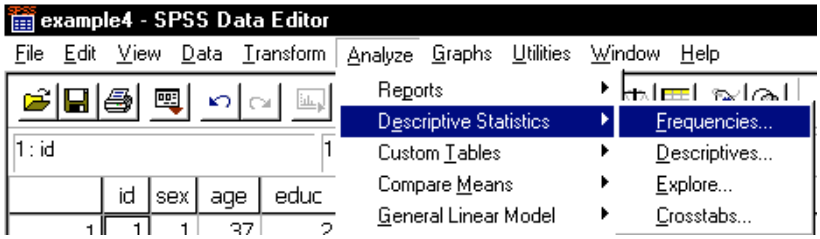
การแจกแจงความถี่จำแนกทางเดียวด้วยโปรแกรม SPSS ด้วยคำสั่ง Analyze / Descriptive Statistics / Frequencies…

ตัวอย่าง 4.11 จงสร้างตารางแจกแจงความถี่ตัวแปร educ ในแฟ้มข้อมูล example4.sav

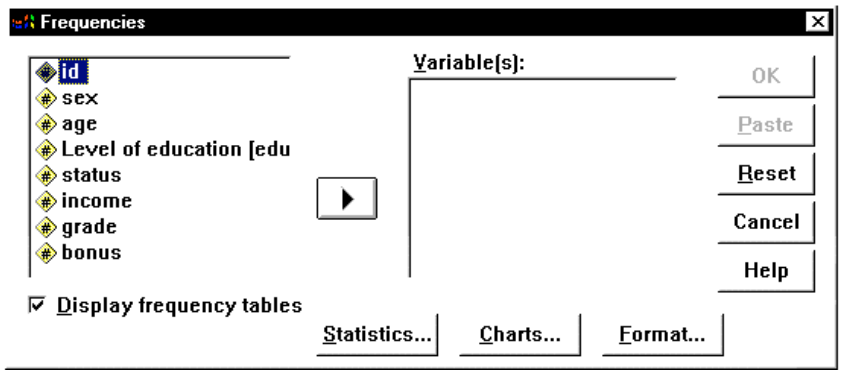
ขั้นที่ 1. นำแฟ้ม example4.sav เข้ามาใน SPSS Data Editor โดยใช้คำสั่ง File / Open



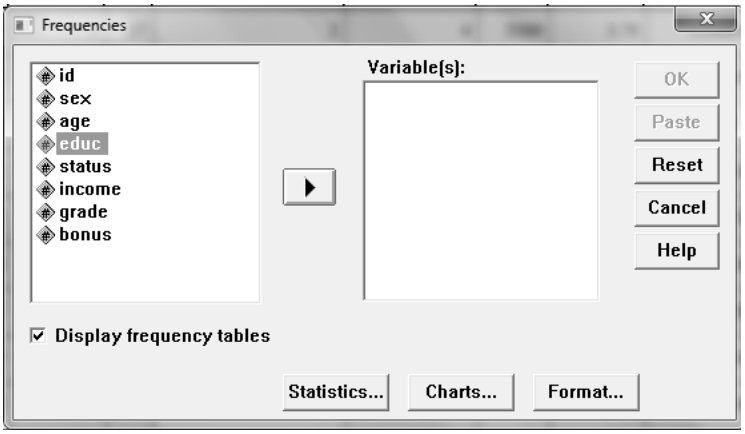
ขั้นที่ 2. เลือกคำสั่ง Analyze / Descriptive Statistics / Frequencies บนจอภาพจะเป็นดังนี้



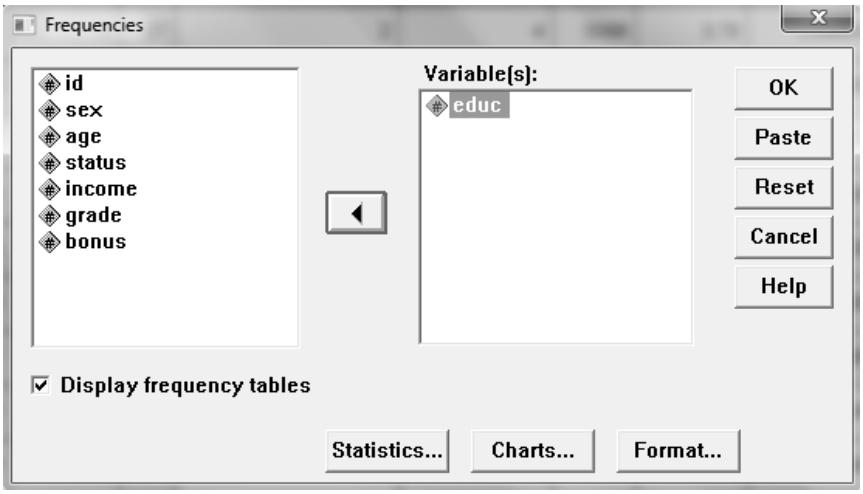
ขั้นที่ 3. คลิกเมาส์ที่ Frequencies จะได้เมนูย่อยของคำสั่งดังนี้



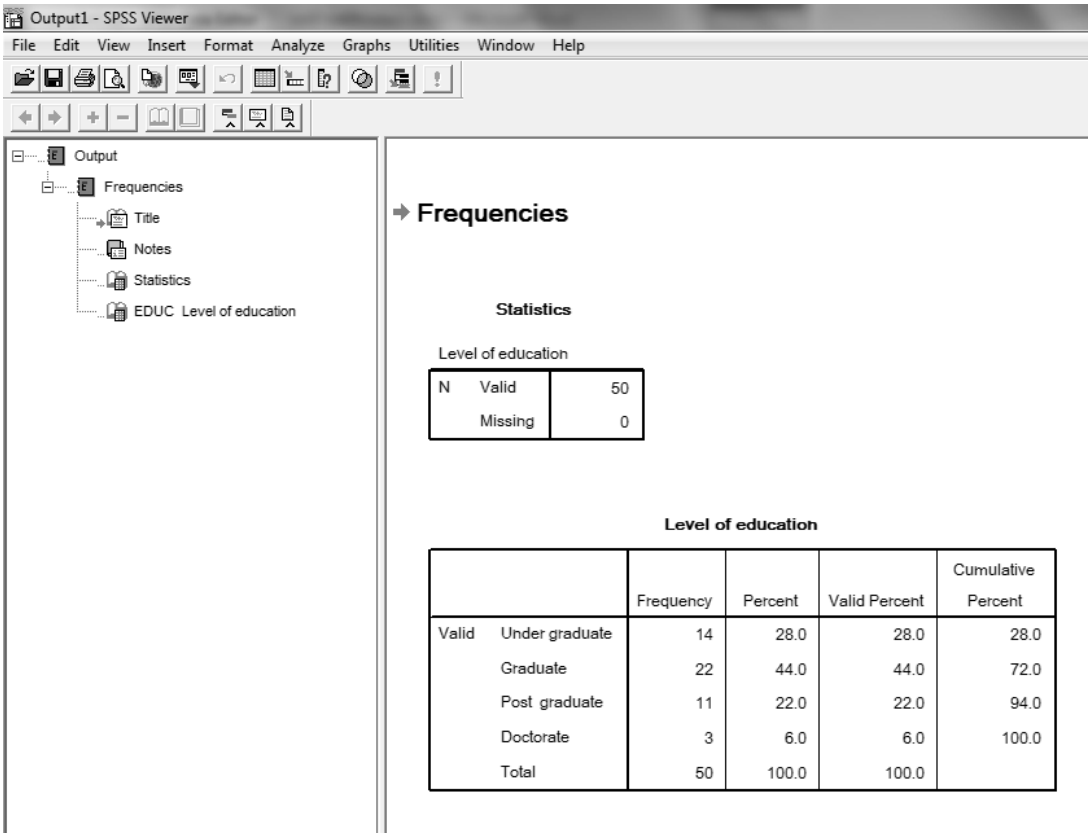
ขั้นที่ 4. เลือกตัวแปรโดยการเอาเมาส์คลิกที่ตัวแปรที่ต้องการ เช่น ตัวแปร educ ขึ้นเป็นแถบสีดำ



ขั้นที่ 5. คลิกที่ปุ่ม ตัวแปร age ที่เราเลือกไว้จะมาอยู่ทางช่องขวามือ



ขั้นที่ 6. คลิก OK ผลการคำนวณที่ SPSS Viewer เป็นดังนี้



ความหมายของผลลัพธ์

Statistics

Level of education

N	Valid	50
	Missing	0

Level of education

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Under graduate	14	28.0	28.0	28.0
	Graduate	22	44.0	44.0	72.0
	Post graduate	11	22.0	22.0	94.0
	Doctorate	3	6.0	6.0	100.0
	Total	50	100.0	100.0	

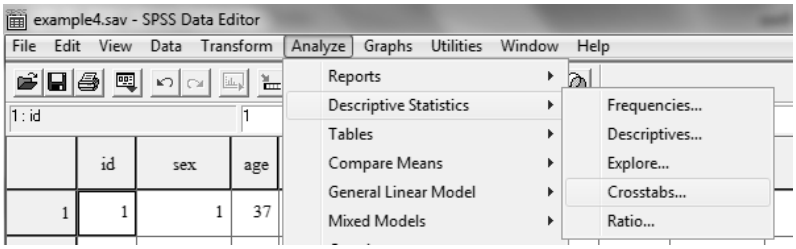
หมายเหตุ การแจกแจงความถี่ของข้อมูลด้วยคำสั่ง Frequencies นั้น ยังสามารถคำนวณค่าสถิติเบื้องต้น และสร้างกราฟได้เช่นกัน แต่เนื่องจากการแจกแจงความถี่เหมาะสมกับข้อมูลเชิงคุณภาพ จึงไม่ขอกล่าวในที่นี้

2. การสร้างตารางแจกแจงความถี่จำแนกสองเดียว

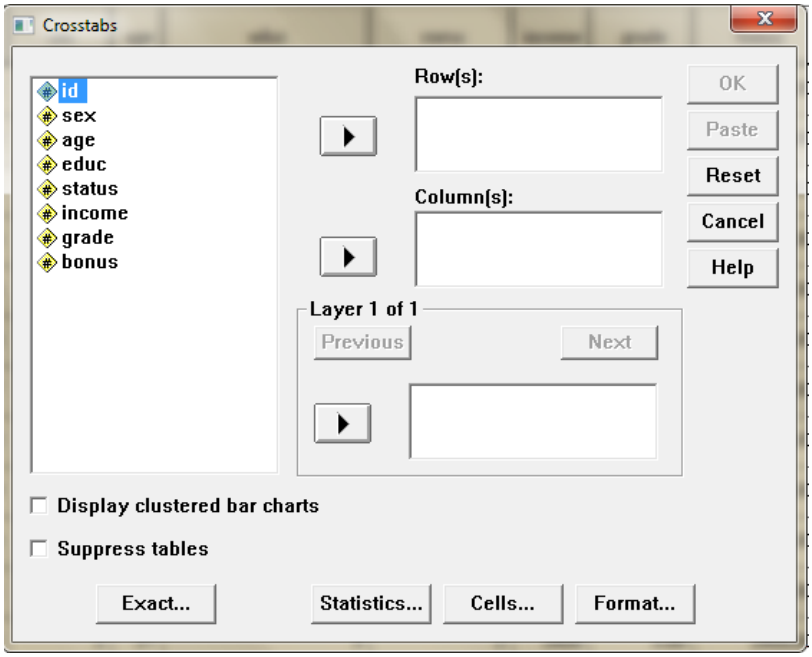
การแจกแจงความถี่จำแนกทางเดียวด้วยโปรแกรม SPSS ด้วยคำสั่ง Analyze / Descriptive Statistics / Crosstabs…

ตัวอย่าง 4.12 จงสร้างตารางแจกแจงความถี่ตัวแปร educ และ status ในแฟ้มข้อมูล example4.sav

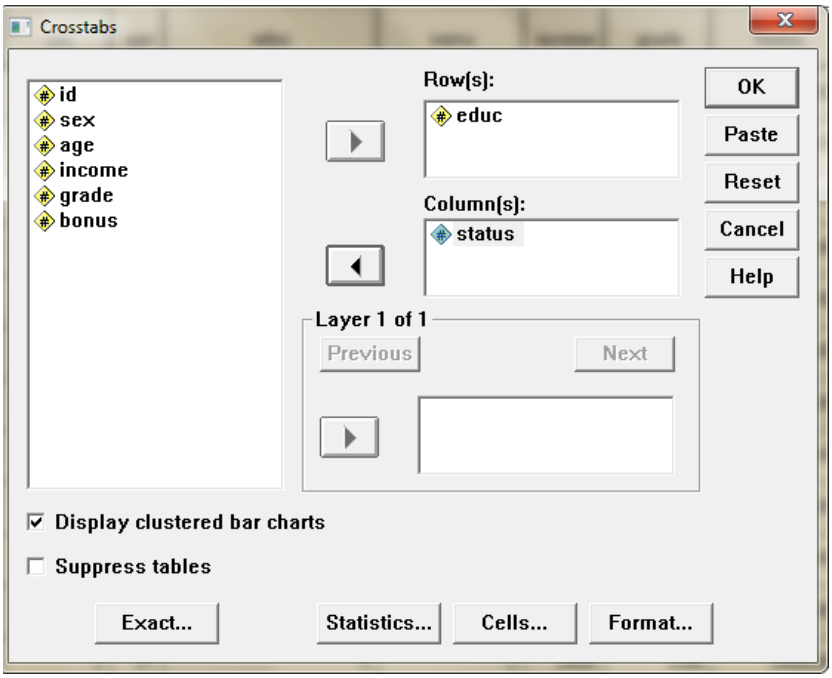
ขั้นที่ 1. เลือกคำสั่ง Analyze / Descriptive Statistics / Crosstabs.. บนจอภาพจะเป็นดังนี้



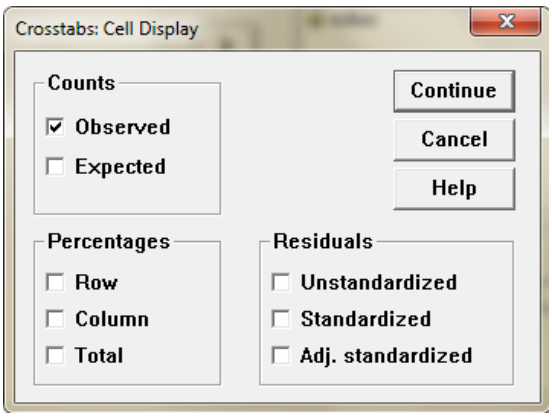
ขั้นที่ 3. คลิกเมาส์ที่ Crosstabs จะได้เมนูย่อยของคำสั่งดังนี้



ขั้นที่ 4 เลือกตัวแปร eudc ไปไว้ในช่อง Row(s) และตัวแปร status ไปไว้ในช่อง Column(s)



หมายเหตุ หากต้องการกำหนดให้ SPSS แสดงค่าร้อยละ ให้คลิกที่ปุ่ม Cells จะได้เมนูย่อย ดังนี้



ในส่วนของ Percentages ให้เลือก Row Column และ Total เลือก Continue -> OK จะได้ผลลัพธ์ดังนี้

Level of education * STATUS Crosstabulation

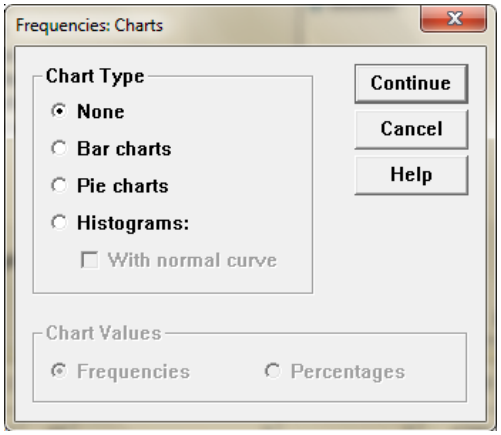
			STATUS				
			Single	Married	Widowhood	Divorce	Total
Level of education	Under graduate	Count	6	6	0	1	13
		% within Level of education	46.2%	46.2%	.0%	7.7%	100.0%
		% within STATUS	42.9%	26.1%	.0%	20.0%	28.3%
		% of Total	13.0%	13.0%	.0%	2.2%	28.3%
	Graduate	Count	3	11	4	3	21
		% within Level of education	14.3%	52.4%	19.0%	14.3%	100.0%
		% within STATUS	21.4%	47.8%	100.0%	60.0%	45.7%
		% of Total	6.5%	23.9%	8.7%	6.5%	45.7%
	Post graduate	Count	4	5	0	1	10
		% within Level of education	40.0%	50.0%	.0%	10.0%	100.0%
		% within STATUS	28.6%	21.7%	.0%	20.0%	21.7%
		% of Total	8.7%	10.9%	.0%	2.2%	21.7%
	Doctorate	Count	1	1	0	0	2
		% within Level of education	50.0%	50.0%	.0%	.0%	100.0%
		% within STATUS	7.1%	4.3%	.0%	.0%	4.3%
		% of Total	2.2%	2.2%	.0%	.0%	4.3%
Total		Count	14	23	4	5	46
		% within Level of education	30.4%	50.0%	8.7%	10.9%	100.0%
		% within STATUS	100.0%	100.0%	100.0%	100.0%	100.0%
		% of Total	30.4%	50.0%	8.7%	10.9%	100.0%

3. การแจกแจงความถี่ด้วยกราฟ

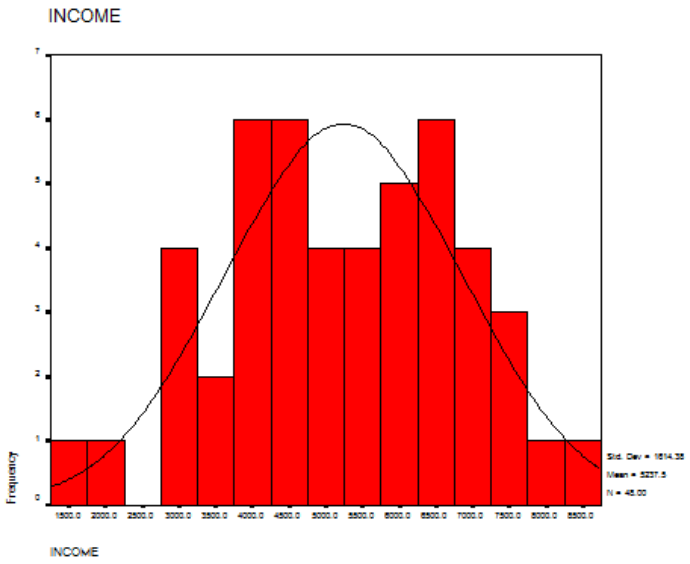
การแจกแจงความถี่ด้วยกราฟทำได้หลายวิธี ในเอกสารฉบับนี้จะกล่าวเฉพาะฮิสโตแกรมเท่านั้น ในโปรแกรม SPSS สามารถสร้างฮีสโตแกรมได้ด้วยคำสั่ง Analyze / Descriptive Statistics / Frequencies…และคำสั่ง Analyze / Descriptive Statistics / Explores…

ตัวอย่าง 4.13 จงสร้างฮีสโตแกรมของตัวแปร income ในแฟ้มข้อมูล example4.sav ด้วยคำสั่ง Analyze / Descriptive Statistics / Frequencies…

ขั้นที่ 1. ดำเนินการตามขั้นที่ 1-3 ในตัวอย่าง 4.11 เลือกตัวแปร incomeไว้ในช่อง Variable(s) เลือกปุ่ม Charts บนจอภาพจะเป็นดังนี้

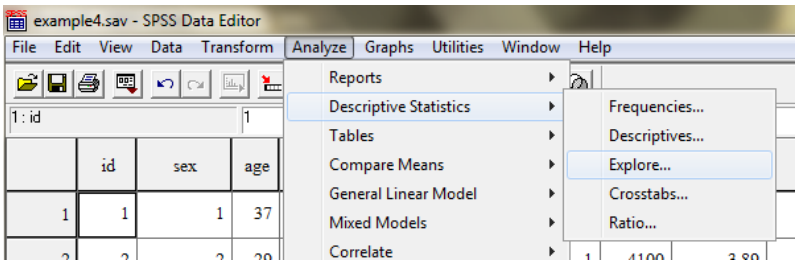


เลือก Histograms และคลิกที่ With normal curve เลือก Continue -> OK ได้ผลลัพธ์ ดังนี้

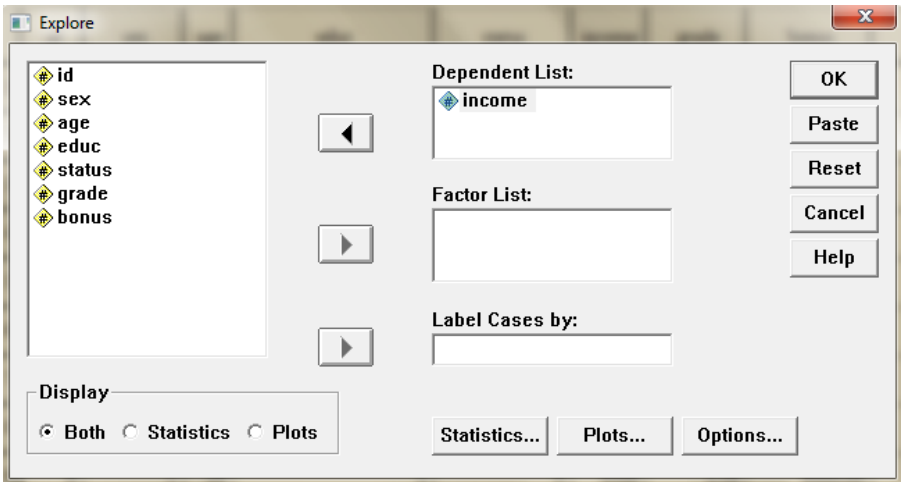


ตัวอย่าง 4.14 จงสร้างฮีสโตแกรมของตัวแปร income ในแฟ้มข้อมูล example4.sav ด้วยคำสั่ง Analyze / Descriptive Statistics / Explore…

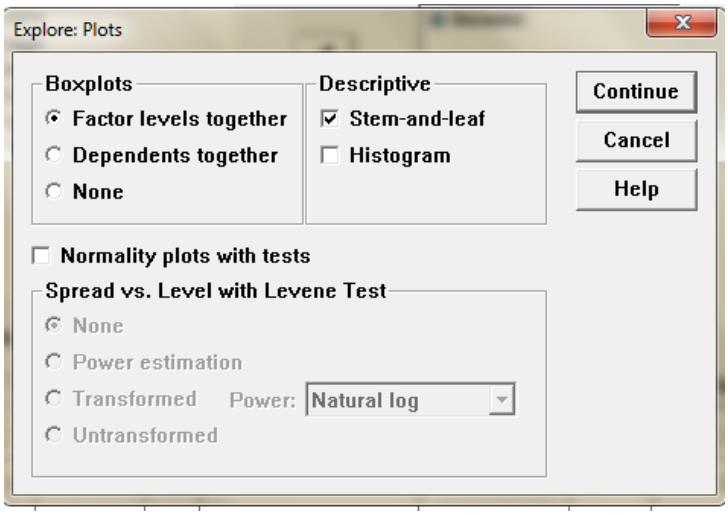
ขั้นที่ 1. เลือกคำสั่ง Analyze / Descriptive Statistics / Explore.. บนจอภาพจะเป็นดังนี้



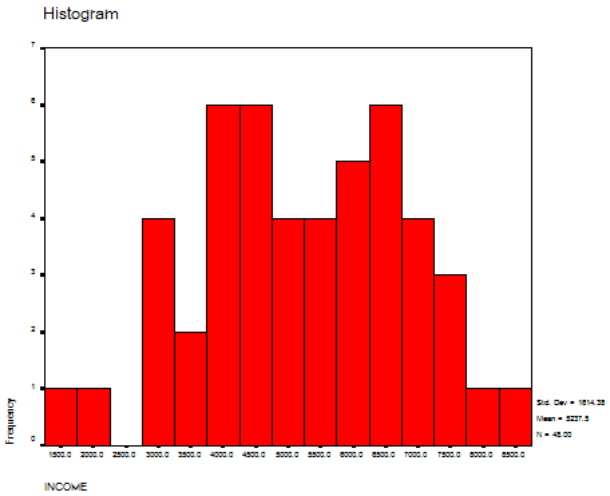
ขั้นที่ 2 เลือกตัวแปร income ไปไว้ในช่อง Dependent List ดังนี้



ขั้นที่ 3 เลือกปุ่ม Polts…บนจอภาพจะเป็นดังนี้



ขั้นที่ 4 เลือก Histogram -> Continue -> OK ได้ผลลัพธ์ดังนี้

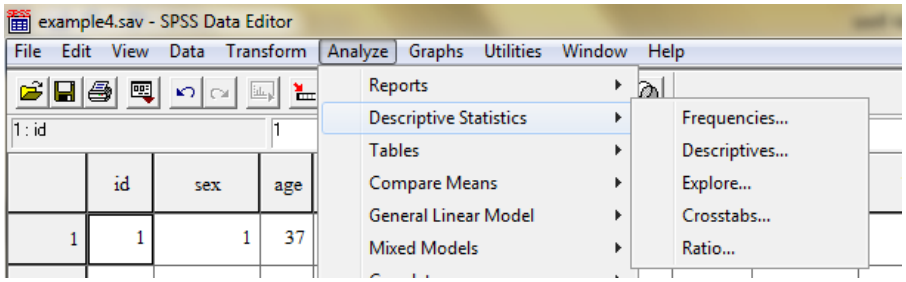


4. การวัดแนวโน้มเข้าสู่ส่วนกลางและการวัดการกระจาย

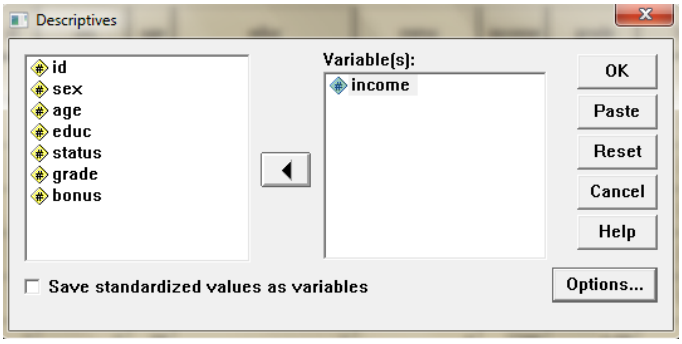
การวัดแนวโน้มเข้าสู่ส่วนกลางและการวัดการกระจายด้วยโปรแกรม SPSS ด้วยคำสั่ง Analyze / Descriptive Statistics / Descriptives…และ Analyze / Descriptive Statistics / Explore…

ตัวอย่าง 4.15 จงวิเคราะห์ข้อมูลเบื้องต้นของตัวแปร income ในแฟ้มข้อมูล example4.sav ด้วยคำสั่ง Analyze / Descriptive Statistics / Descriptives…

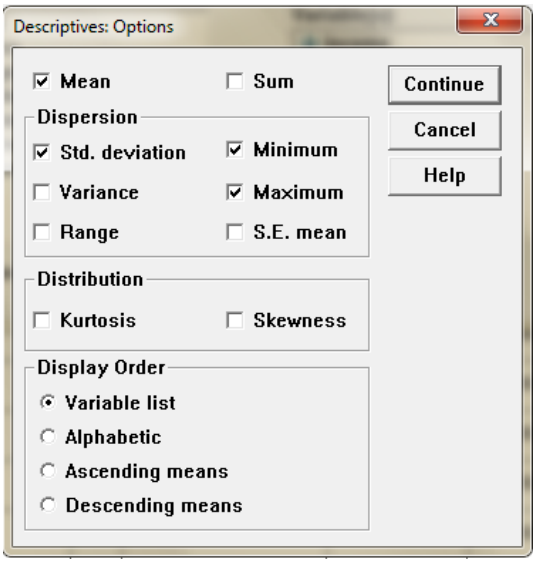
ขั้นที่ 1. เปิดแฟ้มข้อมูล example4.sav เลือกคำสั่ง Analyze / Descriptive Statistics / Descriptives.. บนจอภาพจะเป็นดังนี้



ขั้นที่ 2 เลือกตัวแปร income ไว้ในช่อง Variable(s) ดังนี้



ขั้นที่ 3 ถ้าเลือก OK จะได้ผลลัพธ์เพียงค่าสถิติเบื้องต้นเพียงไม่กี่ค่า หากต้องการค่าสถิติเพิ่มขึ้นให้เลือกปุ่ม Options… บนหน้าจอจะปรากฏผลดังนี้



ขั้นที่ 4 ในส่วน Distribution เลือก Kurtosis และ Skewness เลือก Continue เลือก OK ได้ผลลัพธ์ดังนี้

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std.	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
INCOME	48	1700	8300	5237.50	1614.380	-.136	.343	-.686	.674
Valid N (listwise)	48								

หมายเหตุ หากต้องการแสดงผลลัพธ์ในแนวตั้ง ให้ dubble click ที่ตาราง เลือกเมนู Pivot -> Transpose Rows and Columns จะได้ผลลัพธ์ดังนี้

Descriptive Statistics

		INCOME	Valid N (listwise)
N	Statistic	48	48
Minimum	Statistic	1700	
Maximum	Statistic	8300	
Mean	Statistic	5237.50	
Std. Deviation	Statistic	1614.380	
Skewness	Statistic	-.136	
	Std. Error	.343	
Kurtosis	Statistic	-.686	
	Std. Error	.674	

ตัวอย่าง 4.16 จงวิเคราะห์ข้อมูลเบื้องต้นของตัวแปร income ในแฟ้มข้อมูล example4.sav ด้วยคำสั่ง Analyze / Descriptive Statistics / Explore…

ขั้นที่ 1. เปิดแฟ้มข้อมูล example4.sav เลือกคำสั่ง Analyze / Descriptive Statistics / Explore…เลือกตัวแปร income ไปไว้ในช่อง Dependent List เลือก OK ได้ผลลัพธ์ดังนี้

Descriptives

			Statistic	Std. Error
INCOME	Mean		5237.50	233.016
	95% Confidence Interval for Mean	Lower Bound	4768.73	
		Upper Bound	5706.27	
	5% Trimmed Mean		5259.26	
	Median		5200.00	
	Variance		2606223	
	Std. Deviation		1614.380	
	Minimum		1700	
	Maximum		8300	
	Range		6600	
	Interquartile Range		2675.00	
	Skewness		-.136	.343
	Kurtosis		-.686	.674

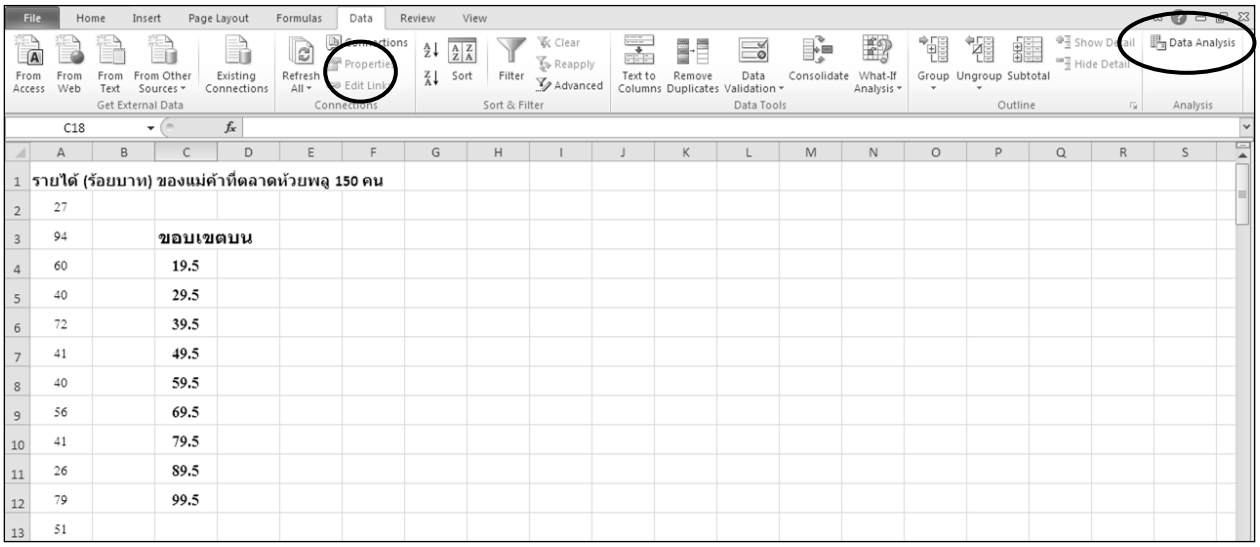
การใช้โปรแกรมสำเร็จรูป MS Excel

การวิเคราะห์ด้วยชุดคำสั่ง Data Analysis

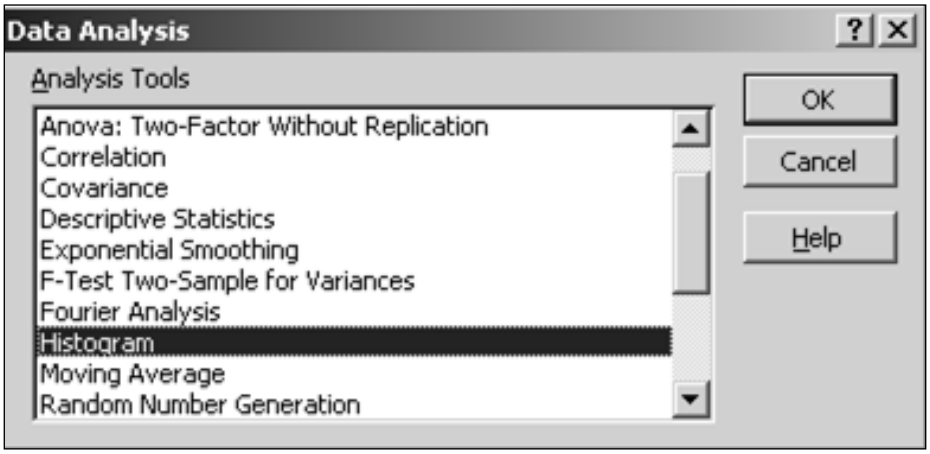
1. การแจกแจงความถี่ด้วยตารางและฮิสโตแกรม

ขั้นที่ 1 ใส่ข้อมูล และขอบเขตบนของตารางแจกแจงความถี่ที่จะสร้าง ไว้ใน cell ที่ต้องการ

เลือกเมนู Data เลือก Data Analysis



ขั้นที่ 2 ในหน้าต่าง Data Analysis เลือก Histogram เลือก OK

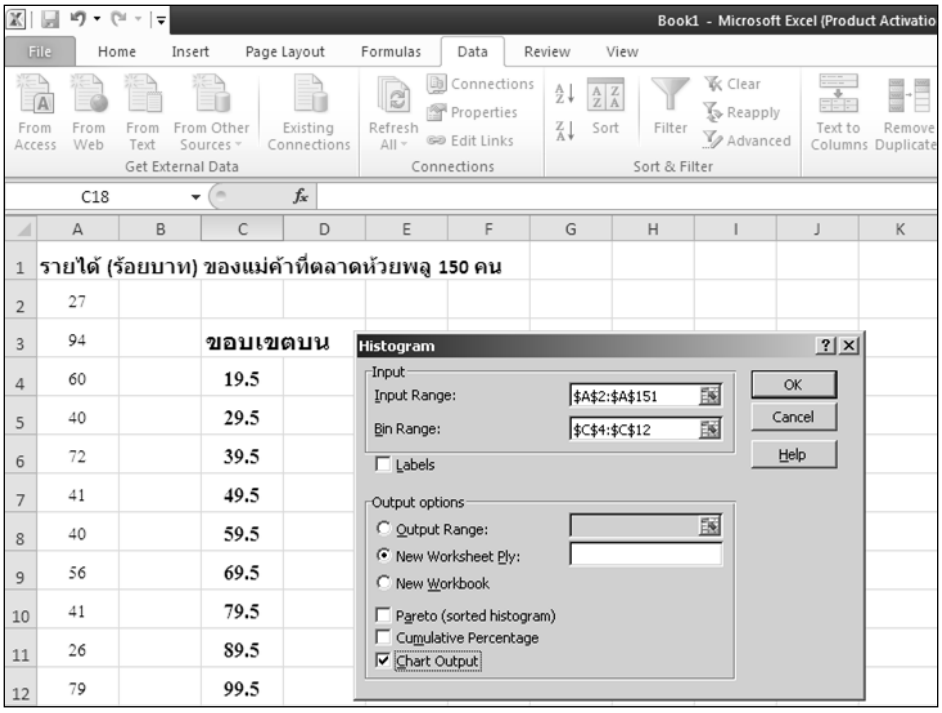


ขั้นที่ 3 ในหน้าต่าง Histogram ในส่วน Input Range ระบุ cell ที่เป็นข้อมูลทั้งหมด

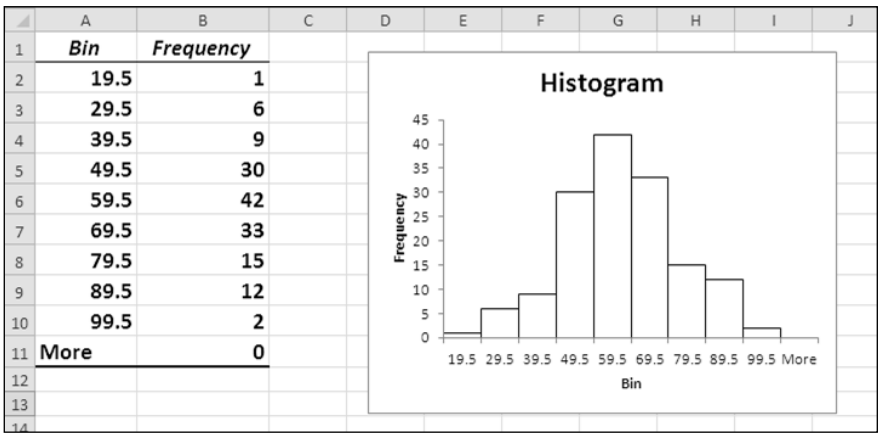
ในส่วน Bin Range ระบุ cell ที่เป็นขอบเขตบน

ในส่วน Output เลือกตำแหน่งผลลัพธ์ และประเภทของแผนภูมิ ดังนี้

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



ขั้นตอนที่ 4 จะได้ผลลัพธ์ ดังนี้

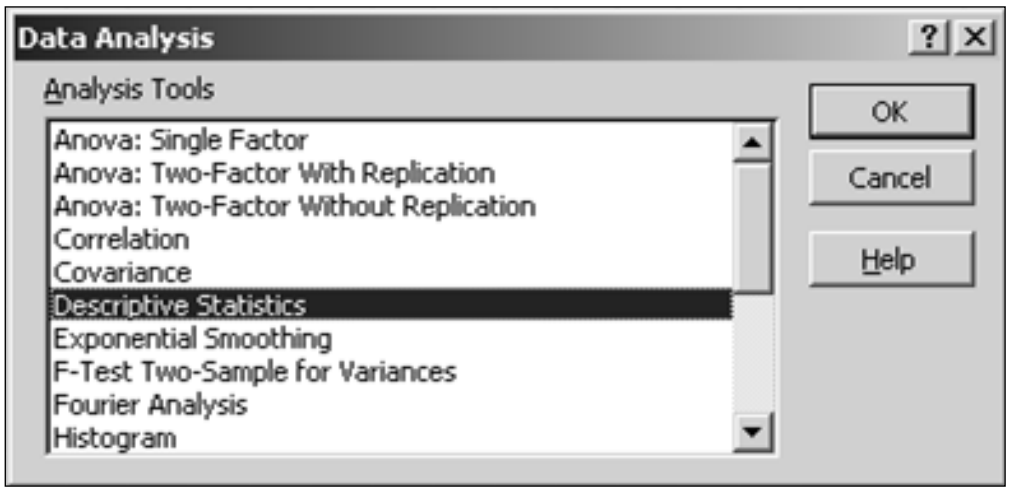


จากผลลัพธ์ตารางแจกแจงความถี่ คอลัมภ์ Bin หมายถึงขอบเขตบนของแต่ละอันตรภาคชั้น เช่นขอบเขตบน 19.5 หมายถึงอันตรภาคชั้น 10-19 ขอบเขตบน 29.5 หมายถึงอันตรภาคชั้น 20-29เป็นต้น คอลัมภ์ Frequency หมายถึงความถี่ของแต่ละอันตรภาคชั้น

จากตารางแจกแจงความถี่พบว่าส่วนใหญ่แม่ค้ามีรายได้ในช่วง 40-69 ร้อยบาท และจากฮิสโตแกรมของรายได้ของแม่ค้าที่ตลาดห้วยพลู 150 คน มีลักษณะคล้ายโค้งสมมาตร แสดงว่ารายได้ของแม่ค้ามีการกระจายเป็นปกติ

2. การวัดแนวโน้มเข้าสู่ส่วนกลาง และการวัดการกระจาย

ขั้นที่ 1 เลือกเมนู Data เลือก Data Analysis เลือก Descriptive Statistics ดังนี้



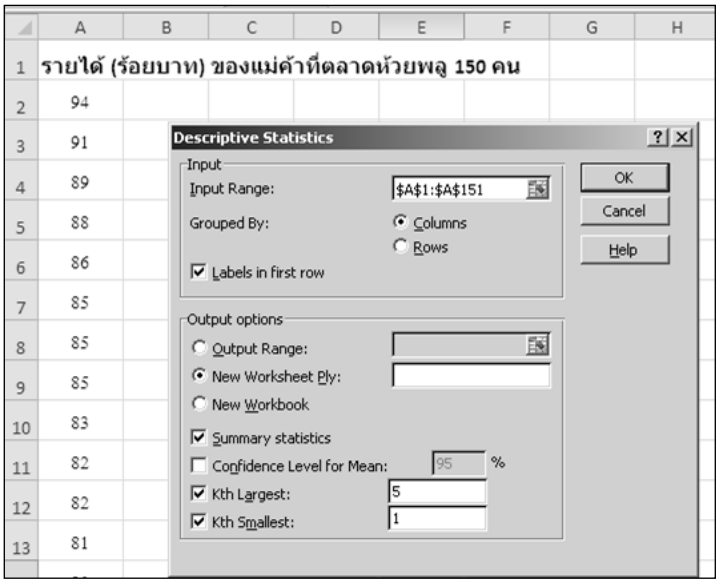
ขั้นที่ 2 ในส่วน Input Range ให้ระบุ cell ที่เป็นข้อมูลทั้งหมด

ในส่วน Group By เป็นการเลือกการแบ่งกลุ่มของข้อมูล

- Columns หมายถึงแบ่งกลุ่มข้อมูลตามคอลัมภ์
- 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



ขั้นที่ 3 จะได้ผลลัพธ์ ดังนี้

	รายได้(ร้อยบาท) ของแม่ค้าที่ตลาดห้วยพลู 150 คน
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**

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 ซึ่งเทคนิคในการวิเคราะห์ข้อมูลเพื่อตอบวัตถุประสงค์เหล่านี้ มีเทคนิคและวิธีการที่แตกต่างกัน 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