

บทที่ 4

การวิเคราะห์ในสถิติพรรณนา

1. แนวคิดเกี่ยวกับสถิติพรรณนา

การวิเคราะห์ข้อมูลสำหรับงานวิจัยทางสังคมศาสตร์มีทั้งวิธีการวิเคราะห์ข้อมูลที่เป็นตัวเลขเชิงปริมาณ (quantitative data) และวิธีการวิเคราะห์ข้อมูลที่มีลักษณะเป็นเชิงคุณภาพ (qualitative data) ในวิธีการวิเคราะห์ข้อมูลทั้ง 2 แบบนี้ก็มีหลายเทคนิควิธีให้เลือกใช้ ซึ่งขึ้นอยู่กับลักษณะของข้อมูล และลักษณะของปัญหาที่ต้องการหาคำตอบในงานวิจัยต่าง ๆ

สถิติพรรณนาเป็นหลักการหนึ่งของวิชาสถิติที่ใช้วิเคราะห์ข้อมูลชุดหนึ่งที่เก็บรวบรวมได้ ผลการวิเคราะห์อยู่ในรูปแบบของการบรรยายเกี่ยวกับคุณลักษณะเบื้องต้นของประชากรที่ศึกษา โดยอาจนำเสนอข้อมูลในรูปตารางแจกแจงความถี่ กราฟ รูปภาพต่าง ๆ หรืออาจนำเสนอค่าวัดคุณลักษณะของสิ่งที่สนใจ โดยส่วนใหญ่ใช้การวัดแนวโน้มเข้าสู่ส่วนกลางและการวัดการกระจาย

การวัดแนวโน้มเข้าสู่ส่วนกลาง วิธีที่ใช้ส่วนใหญ่มีการวัดค่าเฉลี่ยคือ ค่าที่ได้จากการนำข้อมูลทั้งหมดมารวมกัน แล้วหารด้วยจำนวนข้อมูลทั้งหมด มีฐานคือ ค่าที่อยู่ตำแหน่งกลางของข้อมูล และฐานนิยมคือ ค่าที่มีความถี่มากที่สุด

การวัดการกระจายส่วนใหญ่มักจะพิจารณาจากค่าความแปรปรวนและส่วนเบี่ยงเบนมาตรฐานของตัวอย่าง

ในบทนี้จะอธิบายเกี่ยวกับการวิเคราะห์สถิติพรรณนาสำหรับตัวแปรตัวเดียว (univariate descriptive) ซึ่งเป็นเทคนิคพื้นฐานที่ใช้กันทั่วไปสำหรับข้อมูลที่เป็นตัวเลขเชิงปริมาณ ซึ่งต้องอาศัยการคำนวณสูตรต่าง ๆ และการใช้คอมพิวเตอร์ช่วยในการคำนวณสูตรต่าง ๆ เหล่านี้ และการตีความหมายข้อมูลที่ได้จากการวิเคราะห์ในสถิติพรรณนา

2. รูปแบบการนำเสนอข้อมูล

2.1 การแจกแจงความถี่ (Frequency Distributions)

2.1.1 การแจกแจงความถี่สำหรับข้อมูลเชิงคุณภาพ

การตีความหมายข้อมูลโดยการบรรยายเกี่ยวกับคุณลักษณะเบื้องต้นของประชากรที่ศึกษาทำได้โดยเสนอข้อมูลในรูปตารางการแจกแจงความถี่ ในกรณีที่ตัวแปรที่มีลักษณะเป็นเชิงคุณภาพที่มีข้อมูลเป็นจำนวนนับ (count data) อยู่ในแต่ละชั้นของตาราง ซึ่งมีสเกลการวัดข้อมูลแบบแบ่งประเภท (nominal scale) และแบบอันดับ (ordinal scale) ตัวอย่างเช่น การให้ผลสัมฤทธิ์ทางการเรียนวิชาหลักการวางแผนการตลาดของนักเรียน 55 คน แบ่งเป็น 5 ระดับคือ A = เก่งมาก , B = เก่ง , C = ปานกลาง , D = อ่อน และ F = อ่อนมาก ทำการนับจำนวนนักเรียนที่ได้ผลการเรียนเป็นเกรดในระดับต่าง ๆ แสดงเป็นตารางแจกแจงความถี่ในตารางที่ 4.1 ชั้นต่าง ๆ ของตารางคือ ตัวแปรเกรดที่มี 5 ระดับ (A , B , C , D , F) ซึ่งมีสเกลการวัดแบบอันดับ ทำให้เห็นคุณลักษณะของประชากรได้ชัดเจนขึ้นคือ มีนักเรียน 38.2 เปอร์เซ็นต์ของนักเรียนทั้งหมดมีผลสัมฤทธิ์ทางการเรียนดี และมีนักเรียนอ่อน 10.9 เปอร์เซ็นต์

ตารางที่ 4.1 การแจกแจงความถี่ของเกรดวิชาหลักการวางแผนการตลาด

เกรด	ความถี่	เปอร์เซ็นต์
A	1	1.8
B	20	36.4
C	28	50.9
D	6	10.9
F	0	0.0

2.1.2 การแจกแจงความถี่สำหรับข้อมูลเชิงปริมาณ

สำหรับตัวแปรที่มีลักษณะเป็นเชิงปริมาณที่มีข้อมูลเป็นค่าต่อเนื่อง (continuous value) ซึ่งมีสเกลการวัดข้อมูลแบบช่วง (interval scale) และแบบอัตราส่วน (ratio scale) และต้องการบรรยายคุณลักษณะของประชากรในรูปตารางการแจกแจงความถี่ก็ต้องจัดกลุ่มข้อมูลให้เป็นชั้นๆ ก่อน

ตัวอย่างเช่น การพิจารณาผลสัมฤทธิ์ทางการเรียนโดยดูจากคะแนนสอบในรายวิชาหนึ่ง เช่น คะแนนการสอบวิชาหลักการวางแผนการตลาดของนักเรียน 55 คน ในตารางที่ 4.2 ทำความเข้าใจได้ยาก แต่ถ้านำคะแนนสอบนี้มาจัดใหม่ให้อยู่ในรูปตารางการแจกแจงความถี่จะทำให้เห็นชัดเจนกว่าว่ามีนักเรียนที่ได้คะแนนต่าง ๆ จำนวนเท่าไร แต่เนื่องจากข้อมูลคะแนนสอบมีสเกลการวัดแบบช่วง (interval scale) การทำตารางแจกแจงความถี่ของคะแนนสอบก็ทำได้โดยที่บางค่าของคะแนนสอบอาจไม่มีเลยก็ได้ ดังตารางที่ 4.3

ตารางที่ 4.2 คะแนนสอบวิชาหลักการวางแผนการตลาดของนักเรียน 55 คน

22	27	64	50	56	59
23	49	39	46	41	49
27	52	44	40	51	45
33	55	50	50	48	41
47	52	56	37	51	55
56	36	39	41	44	
39	48	42	54	41	
53	47	47	54	46	
54	45	48	25	50	
46	48	56	35	53	

ตารางที่ 4.3 การแจกแจงความถี่ของคะแนนสอบวิชาหลักการวางแผนการตลาด

คะแนน	ความถี่ (frequency)	เปอร์เซ็นต์ %	เปอร์เซ็นต์ สะสม
22	1	1.8	1.8
23	1	1.8	3.6
25	1	1.8	5.4
27	2	3.6	9.0
33	1	1.8	10.8
35	1	1.8	12.6
36	1	1.8	14.4
37	1	1.8	16.2
39	3	5.5	21.7
40	1	1.8	23.5
41	4	7.3	30.8
42	1	1.8	32.6
44	2	3.6	39.8
45	2	3.6	45.3
46	3	5.5	50.8
47	3	5.5	58.1
48	4	7.3	61.7
49	2	3.6	69.0
50	4	7.3	72.6
51	2	3.6	76.2
52	2	3.6	79.8
53	2	3.6	85.3
54	3	5.5	88.9
55	2	3.6	96.2
56	4	7.3	98.0
59	1	1.8	100.0
64	1	1.8	
ผลรวม	55	100	

การจัดกลุ่มคะแนนสอบให้เป็นชั้น ๆ โดยให้แต่ละชั้นมีช่วงคะแนนเท่ากับ 5 คะแนน เป็นการจัดข้อมูลที่มีสเกลการวัดแบบช่วงให้เป็นชั้น ๆ (categorical data) ก่อน แล้วจึงจัดทำเป็นตารางแจกแจงความถี่ซึ่งทำให้การตีความหมายข้อมูลได้ชัดเจนขึ้น ดังตารางที่ 4.4

ตารางที่ 4.4 การแจกแจงความถี่ของคะแนนสอบที่แบ่งออกเป็น 9 ชั้น

คะแนน	ความถี่	เปอร์เซ็นต์
20 – 24	2	3.6
25 – 29	3	5.5
30 – 34	1	1.8
35 – 39	6	10.9
40 – 44	8	14.6
45 – 49	14	25.5
50 – 54	13	23.6
55 – 59	7	12.7
60 – 64	1	1.8

การจัดกลุ่มข้อมูลให้เป็นชั้น ๆ นี้มีประโยชน์สำหรับการวิเคราะห์ข้อมูลแบบแบ่งชั้น (analysis of categorical data) ซึ่งจะอธิบายในบทอื่นต่อไป

2.2 สัดส่วนและเปอร์เซ็นต์ (proportions and percentages)

ต่อจากนั้นเราสามารถตีความหมายข้อมูลเพื่อให้เห็นขนาดของข้อมูล โดยการคำนวณค่า สัดส่วน (proportion) แทนด้วย p และเปอร์เซ็นต์ (percentage) แทนด้วย % มีสูตรการคำนวณคือ

$$p = \frac{f}{n}$$

และ
$$\% = \frac{f}{n} \times 100$$

เมื่อ f คือ จำนวนความถี่ในชั้นหนึ่ง ๆ ของตารางการแจกแจงความถี่

n คือ ขนาดของตัวอย่าง

เราใช้เปอร์เซ็นต์ในการเปรียบเทียบจำนวนความถี่ในชั้นใดชั้นหนึ่งของตารางกับขนาดของตัวอย่างได้เป็นขนาดของข้อมูลในชั้นใดชั้นหนึ่งนั้นเทียบกับจำนวนข้อมูลทั้งหมด โดยมีฐานเป็น 100 และยังสามารถใช้วัดความเปลี่ยนแปลงของค่าของสิ่งที่สนใจเมื่อเวลาผ่านไปช่วงระยะเวลาหนึ่ง ทำให้เห็นการเปลี่ยนแปลงได้ เรียกว่า **Percentage change** มีสูตรการคำนวณคือ

$$\text{Percentage change} = \frac{(\text{จำนวนที่จุดเวลา 2}) - (\text{จำนวนที่จุดเวลา 1})}{\text{จำนวนที่จุดเวลา 1}} \times 100$$

ตัวอย่างเช่น ในปี พ.ศ. 2500 เมือง A มีประชากรประมาณ 2 ล้านคน และเมือง B มีประชากรประมาณ 1 ล้านคน ต่อมาในปี พ.ศ. 2545 เมือง A มีประชากร 14 ล้านคน และเมือง B มีประชากร 4 ล้านคน จะเห็นว่าเมือง A มีประชากรเพิ่มขึ้น 12 ล้านคน และเมือง B มีประชากรเพิ่มขึ้น 3 ล้านคน เนื่องจากที่จุดเวลาเริ่มต้นเมือง A และ B มีขนาดของประชากรแตกต่างกัน อันที่จริงเราอยากทราบเปอร์เซ็นต์ที่เพิ่มขึ้นของประชากรในทั้ง 2 กรณีนี้ การคำนวณเปอร์เซ็นต์ที่เปลี่ยนแปลงสำหรับเมือง A และ B คือ

$$\text{เมือง A : Percentage change} = \frac{14 - 2}{2} \times 100 = 85.7\%$$

$$\text{เมือง B : Percentage change} = \frac{4 - 1}{1} \times 100 = 75.0\%$$

พิจารณาจาก Percentage change ทำให้เห็นการเพิ่มขึ้นของประชากรได้ชัดเจนว่าเมือง A มีการเพิ่มขึ้นของประชากรมากกว่าเมือง B

2.3 อัตราส่วน (Ratios)

อัตราส่วนเป็นการเปรียบเทียบจำนวน 2 จำนวน ที่อยู่ในชั้นของตาราง 2 ชั้น มีสูตรการคำนวณคือ

$$\text{Ratio} = \frac{\text{จำนวนในชั้นที่มีข้อมูลมากกว่า}}{\text{จำนวนในชั้นที่มีข้อมูลน้อยกว่า}}$$

ตัวอย่างเช่น นักเรียนที่เรียนวิชาหลักการวางแผนการตลาดเป็นผู้ชาย 10 คน และผู้หญิง 45 คน คำนวณหาอัตราส่วนของเพศชายต่อเพศหญิงคือ

$$\text{Ratio} = \frac{45}{10} = 4.5$$

อัตราส่วนของเพศชายต่อเพศหญิงคือ 1 : 4.5 นั่นคือสำหรับผู้ชายทุก ๆ 1 คน มีผู้หญิง 4.5 คน

เราสามารถใช้อัตราส่วนในการเปรียบเทียบจำนวนมากกว่า 2 จำนวนก็ได้

ตัวอย่างเช่น ต้องการหาอัตราส่วนของจำนวนผู้ที่ได้เกรดต่าง ๆ คือ เกรด D ต่อเกรด C ต่อเกรด B ต่อเกรด A วิธีการคือให้จำนวนผู้ที่ได้เกรด D เป็นฐานคือ 1 อาศัยข้อมูลจำนวนผู้ที่ได้เกรดต่าง ๆ ในตารางที่ 4.1

อัตราส่วนของเกรด D : เกรด C : เกรด B : เกรด A คือ

$$\frac{6}{6} : \frac{28}{6} : \frac{20}{6} : \frac{1}{6} \text{ เท่ากับ } 1 : 4.7 : 3.3 : 0.2$$

2.4 อัตรา (Rates)

อัตราเป็นค่าความผันแปรโดยคิดจากการแปลงค่าของข้อมูลที่เป็นตัวเลขจำนวนมาก ให้เป็นตัวเลขจำนวนเล็กลงโดยมีเลขฐาน เป็น 100 , 1,000 , หรือ 100,000 เป็นต้น ใช้สำหรับเป็นค่าสรุปแบบรวมและเปรียบเทียบเหตุการณ์ที่เกิดขึ้นในประชากรหนึ่งหรือในชั้นหนึ่งกับช่วงเวลาอื่น ๆ หรือเปรียบเทียบระหว่างประชากรที่แตกต่างกันหรืออยู่ในชั้นต่าง ๆ เช่น อัตราการเกิด อัตราการตาย อัตราการเกิดอุบัติเหตุในท้องถนน อัตราการฆ่าตัวตาย เป็นต้น ค่าของอัตราคล้ายกับอัตราส่วน สัดส่วน หรือเปอร์เซ็นต์ แต่ค่าของอัตรามีฐานที่ใหญ่กว่าโดยคูณกับจำนวนที่ใช้กำหนดให้เป็นตัวเลขฐาน ตัวอย่างเช่น ถ้าให้ฐานเป็น 1,000 สูตรการคำนวณคือ

$$\text{อัตรา} = \frac{\text{จำนวนนับของการเกิดเหตุการณ์ที่สนใจ}}{\text{จำนวนประชากรทั้งหมดที่เกิดเหตุการณ์ต่างๆ}} \times 1,000$$

2.5 กราฟรูปภาพ

การแสดงการแจกแจงความถี่ด้วยกราฟรูปภาพที่สำคัญมี 4 แบบคือ

2.5.1 กราฟแท่ง (Bar Charts)

ใช้สำหรับแสดงความถี่หรือเปอร์เซ็นต์ของข้อมูลที่มีสเกลการวัดแบบแบ่งประเภทและแบบอันดับ ทำให้เห็นคุณลักษณะของประชากรคือ ขนาดของข้อมูลในแต่ละชั้นได้ชัดเจน ตัวอย่างเช่น ภาพที่ 4.3 แสดงกราฟแท่งของความถี่ของข้อมูลเกรดวิชาหลักการวางแผนการตลาดในตารางที่ 4.1

เราสามารถโปรแกรม SPSS ช่วยในการสร้างกราฟได้ โดยดำเนินการตามขั้นตอนดังนี้

1. สร้างแฟ้มข้อมูลคะแนนสอบวิชาหลักการวางแผนการตลาด โดยกำหนดให้คะแนนสอบแทนด้วยตัวแปร Score ซึ่งมีสเกลการวัดแบบช่วง และเกรดแทนด้วยตัวแปร Grade ซึ่งได้จากการจัดกลุ่มคะแนนเป็น 4 กลุ่ม คือ D = อ่อน , C = ปานกลาง , B = ดี , และ A = ดีมาก บันทึกข้อมูลในแฟ้มข้อมูลชื่อ Subject1.sav มีรูปแบบของข้อมูลดังแสดงในตารางที่ 4.5

ตารางที่ 4.5 รูปแบบข้อมูลคะแนนสอบและเกรดในแฟ้มข้อมูล Subject1.sav

Score	Grade	Score	Grade	Score	Grade	Score	Grade
22	D	41	C	48	C	54	B
23	D	41	C	48	C	54	B
25	D	42	C	49	C	55	B
27	D	44	C	49	C	55	B
27	D	44	C	50	B	56	B
33	D	45	C	50	B	56	B
35	C	45	C	50	B	56	B
36	C	46	C	50	B	56	B
37	C	46	C	51	B	59	B
39	C	46	C	51	B	64	A
39	C	47	C	52	B		
39	C	47	C	52	B		
40	C	47	C	53	B		
41	C	48	C	53	B		
41	C	48	C	54	B		

2. สร้างกราฟแท่ง มีวิธีการคือ

1. ไปที่เมนูบาร์ คลิกที่ Graphs, Bar ... จะได้นหน้าต่าง Bar Charts ดังภาพที่ 4.1

2. ในหน้าต่าง Bar Charts

- เลือกที่ กราฟแท่ง Simple

ในกรอบ Data in Chart Are

- เลือก Summaries for groups of cases

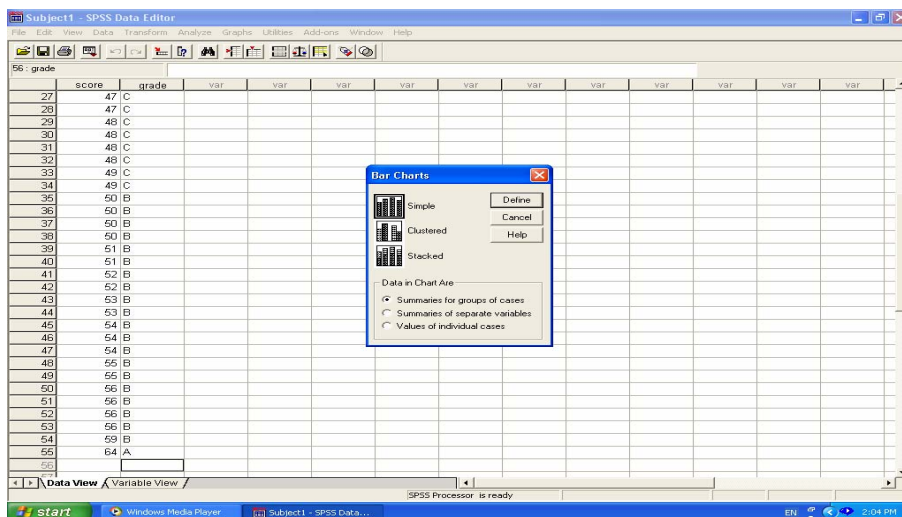
แล้วคลิกที่ปุ่ม Define จะได้นหน้าต่าง Define Simple Bar : Summaries for Groups of Cases ดังภาพที่ 4.2

3. ในหน้าต่าง Define Simple Bar : Summaries for Groups of Cases

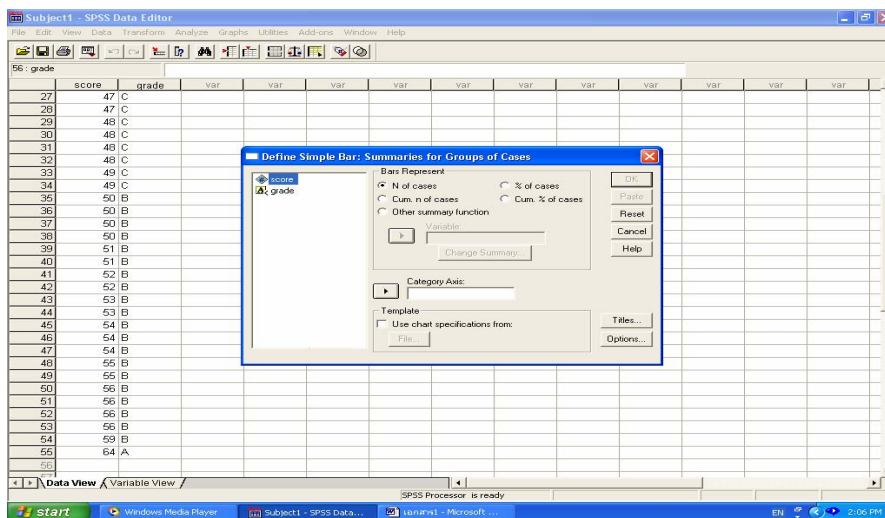
ในช่องซ้ายมือ คลิกเลือกตัวแปร Grade ให้ย้ายเข้าไปอยู่ในช่อง Category Axis:

ในกรอบ Bar Represent เลือก N of cases

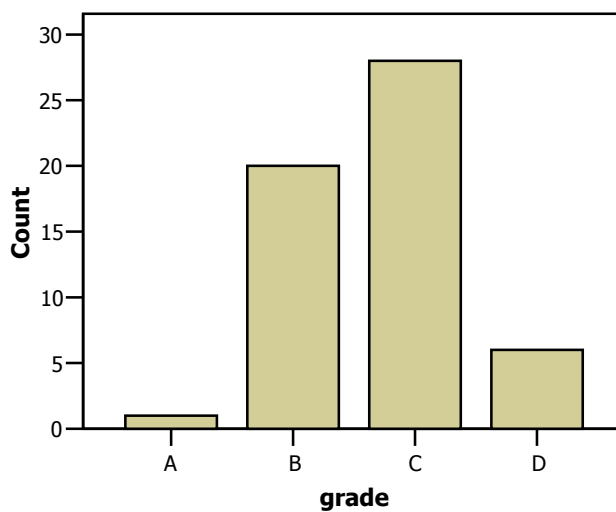
แล้วคลิกที่ปุ่ม OK จะได้ผลลัพธ์ดังภาพที่ 4.3



ภาพที่ 4.1



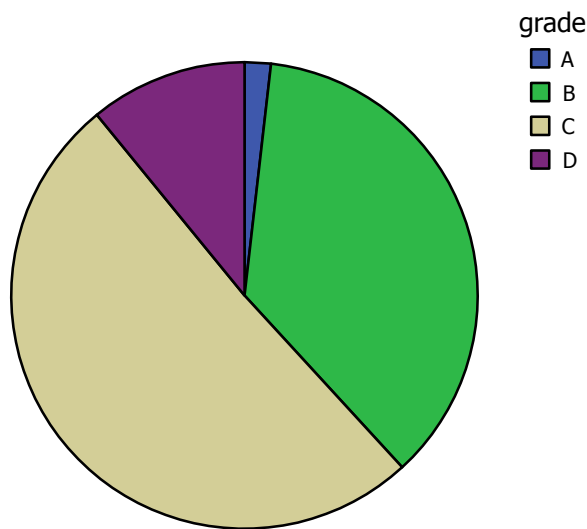
ภาพที่ 4.2



ภาพที่ 4.3 กราฟแท่งของข้อมูลเกรดวิชาหลักการวางแผนการตลาด

2.5.2 กราฟวงกลม (Pie Charts)

ใช้แสดงความถี่หรือเปอร์เซ็นต์ของข้อมูลที่มีสเกลการวัดแบบแบ่งประเภทและแบบอันดับ เช่นเดียวกับกราฟแท่ง เพียงแต่เปลี่ยนจากแท่ง (bar) เป็นเซ็กเมนต์ (segments) ของวงกลม โดยให้พื้นที่ของเซ็กเมนต์หนึ่งเท่ากับความถี่หรือเปอร์เซ็นต์ของข้อมูลในชั้นนั้น ตัวอย่างเช่น การสร้างกราฟวงกลมของความถี่ของข้อมูลเกรดในตารางที่ 4.5 เราสามารถใช้โปรแกรม SPSS ช่วยในการสร้างกราฟได้เช่นเดียวกับการสร้างกราฟแท่งที่อธิบายไว้แล้ว แตกต่างกันเพียงการใช้คำสั่งเลือกชนิดของกราฟ คือ ใช้คำสั่ง Graphs , Pie ... จะได้ผลลัพธ์ดังภาพที่ 4.4



ภาพที่ 4.4 กราฟวงกลมของข้อมูลเกรดวิชาหลักการวางแผนการตลาด

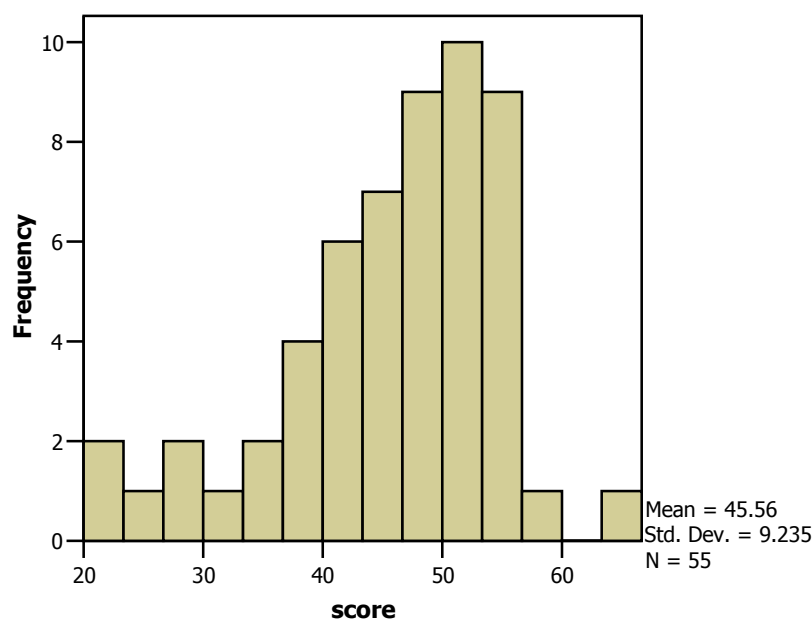
2.5.3 ฮิสโตแกรม (Histogram)

สำหรับข้อมูลที่เป็นตัวเลขเชิงปริมาณที่มีสเกลการวัดเป็นตัวเลขแบบต่อเนื่อง สามารถใช้วิธีการจัดกลุ่มข้อมูลให้เป็นชั้นก่อน แล้วแสดงเป็นกราฟรูปภาพแบบฮิสโตแกรม หรือกราฟเส้นก็ได้ ฮิสโตแกรมคือ กราฟแท่งที่มีแท่งติดกันตลอดไม่มีช่องว่างระหว่างแท่งซึ่งเป็นไปตามลักษณะของข้อมูลที่มีสเกลการวัดเป็นตัวเลขแบบต่อเนื่อง โดยที่ความสูงของแท่งแสดงความถี่ของข้อมูล และความกว้างของแท่งแทนความกว้างของชั้นของข้อมูล พื้นที่ของแท่งคือ

ความถี่สัมพัทธ์ (relative frequency) ของข้อมูลในชั้นหนึ่ง ๆ ทำให้การนำเสนอข้อมูลมีความถูกต้องมากขึ้น ตัวอย่างเช่น การสร้างฮิสโตแกรมของข้อมูลคะแนนสอบวิชาหลักการวางแผนการตลาดของนักเรียน 55 คน ข้อมูลอยู่ในตารางที่ 4.5 เราสามารถใช้โปรแกรม SPSS ช่วยในการสร้างฮิสโตแกรม คือ

1. ไปที่เมนูบาร์เลือก Graphs , Histogram... จะได้นหน้าต่าง Histogram
2. ในหน้าต่าง Histogram

ในช่องซ้ายมือ เลือกตัวแปร score ให้ย้ายเข้าไปอยู่ในช่อง Variable : แล้วคลิกที่ปุ่ม OK จะได้ผลลัพธ์ดังภาพที่ 4.5



ภาพที่ 4.5 ฮิสโตแกรมของข้อมูลคะแนนสอบวิชาหลักการวางแผนการตลาด

2.5.4 กราฟเส้น (Line Graph)

ฮิสโตแกรมสามารถแปลงให้เป็นกราฟเส้นได้โดยการลากเส้นเชื่อมต่อระหว่างจุดกลางของแต่ละแท่งในฮิสโตแกรม ยังมีจำนวนชั้นมากขึ้นเพียงใดกราฟเส้นที่ได้จะเข้าใกล้เส้นโค้งมากขึ้นเพียงนั้น ตัวอย่างเช่น การสร้างกราฟเส้นของข้อมูลเกรดวิชาหลักการวางแผนการตลาดของนักเรียน 55 คน ข้อมูลอยู่ในตารางที่ 4.5 เราสามารถใช้โปรแกรม SPSS ช่วยในการสร้าง คือ

1. ไปที่เมนูบาร์ เลือก Graphs , Line... จะได้นหน้าต่าง Line Charts ดังภาพที่ 4.6

2. ในหน้าต่าง Line Charts

เลือกภาพ กราฟเส้น Simple

แล้วคลิกที่ปุ่ม Define จะได้นหน้าต่าง Define Simple Line : Summaries for Groups of Cases

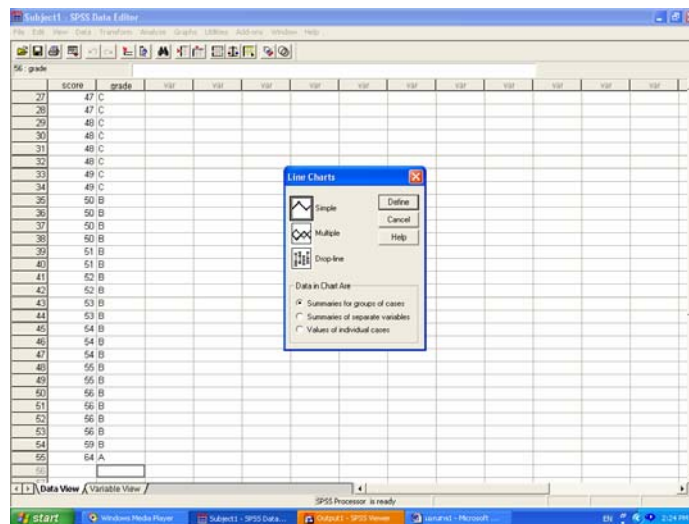
3. ในหน้าต่าง Define Simple Line : Summaries for Groups of Cases

ในช่องซ้ายมือ คลิกเลือกตัวแปร Grade ให้ย้ายเข้าไปอยู่ในช่อง Category

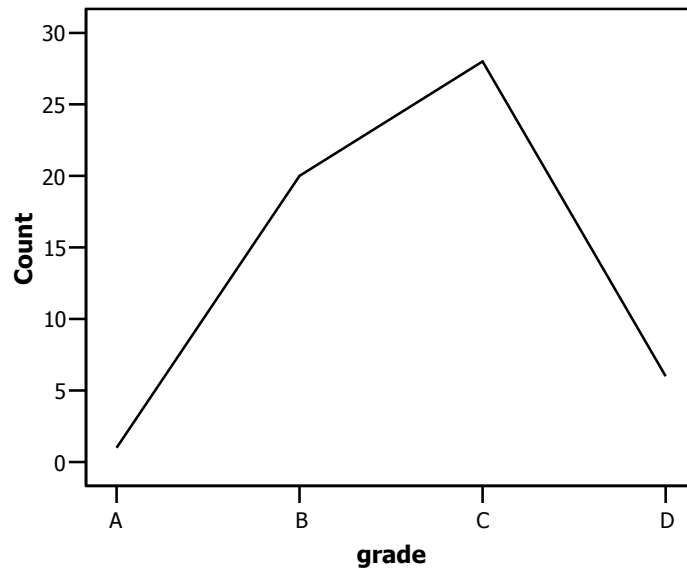
Axis:

ในกรอบ Line Represents เลือก O N of cases

แล้วคลิกที่ปุ่ม OK จะได้ผลลัพธ์ดังภาพที่ 4.7



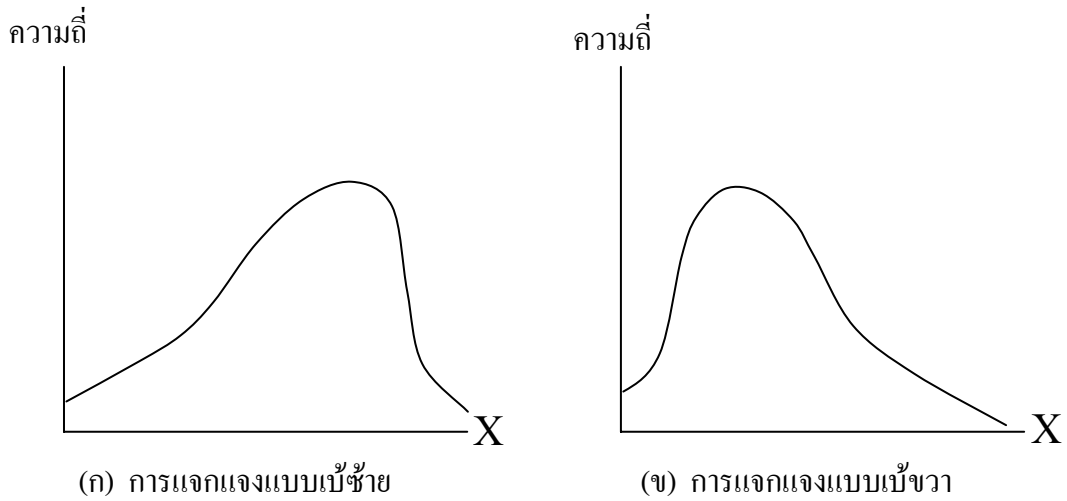
ภาพที่ 4.6



ภาพที่ 4.7 กราฟเส้นของข้อมูลเกรดวิชาหลักการวางแผนการตลาด

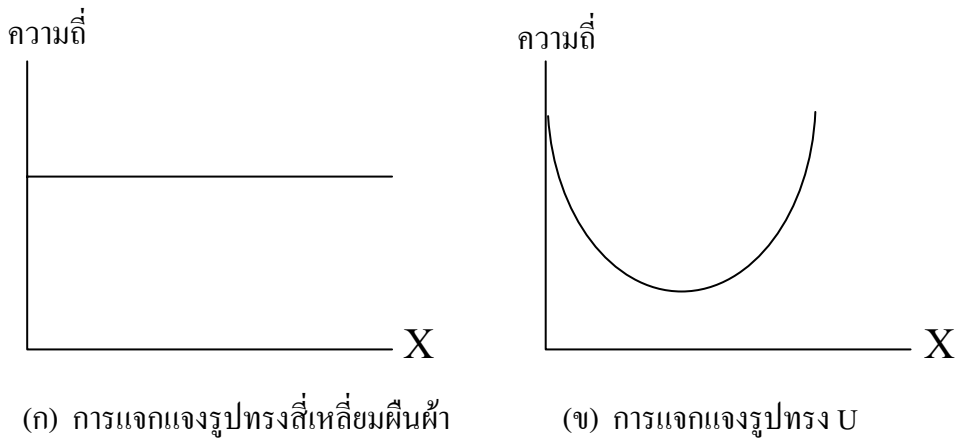
2.6 รูปแบบของการแจกแจงความถี่

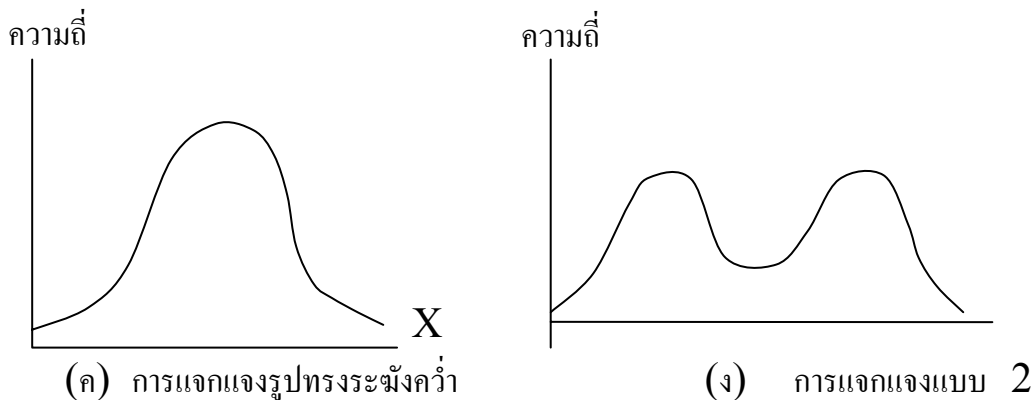
สำหรับการแจกแจงความถี่ของตัวแปรที่มีสเกลการวัดแบบต่อเนื่อง มีรูปทรงแตกต่างกันได้หลายแบบอาจเป็นแบบสมมาตรหรือเบ้ก็ได้ ถ้าพีคของการแจกแจงที่ตรงจุดกึ่งกลางตามแนวแกนตั้งจะแบ่งพื้นที่ใต้โค้งออกได้เป็น 2 ส่วนที่ทับกันสนิท เรียกว่า การแจกแจงสมมาตร (symmetrical distribution) แต่ถ้าไม่เท่ากันหรือไม่ทับกันสนิท เรียกว่า การแจกแจงแบบเบ้ (skewed distributions) อาจเป็นแบบเบ้ซ้าย หรือเบ้ขวา ตัวอย่างเช่น ภาพที่ 4.8 (ก) แสดงการแจกแจงแบบเบ้ซ้าย (negatively skewed) และภาพที่ 4.8 (ข) แสดงการแจกแจงแบบเบ้ขวา (positively skewed)



ภาพที่ 4.8 การแจกแจงแบบเบ้

สำหรับการแจกแจงแบบสมมาตรมีได้หลายรูปทรง ตัวอย่างเช่น ภาพที่ 4.9 (ก) แสดงการแจกแจงรูปทรงสี่เหลี่ยมผืนผ้า (Rectangular distribution) ภาพที่ 4.9 (ข) แสดงการแจกแจงรูปทรง U (U – shaped distribution) ภาพที่ 4.9 (ค) แสดงการแจกแจงรูปทรงระฆังคว่ำ (bell-shaped distribution) และภาพที่ 4.9 (ง) แสดงการแจกแจงแบบ 2 ยอด (bimodal distribution)





ยอด

ภาพที่ 4.9 การแจกแจงแบบสมมาตร

การแจกแจงรูปทรงระฆังคว่ำที่สำคัญมากในทฤษฎีทางสถิติคือ การแจกแจงแบบปกติ (normal distribution) ซึ่งใช้เป็นทฤษฎีพื้นฐานในการประมาณค่าพารามิเตอร์ของประชากรจากค่าสถิติของกลุ่มตัวอย่าง และมักอยู่ในข้อตกลงเบื้องต้นของการใช้สถิติวิเคราะห์หลายวิธีซึ่งจะเห็นได้ในบทต่อ ๆ ไป

2.7 การวัดแนวโน้มเข้าสู่ส่วนกลาง (measures of central tendency)

การวัดแนวโน้มเข้าสู่ส่วนกลางคือ การหาค่าของข้อมูลที่เป็นค่ากลางหรืออยู่ตรงตำแหน่งที่เป็นจุดกลางหรือจุดที่มากที่สุดในการแจกแจงหนึ่ง การวัดค่ากลางที่ใช้กันทั่วไปมี 3 ค่าคือ ค่าฐานนิยม (mode) ค่ามัธยฐาน (median) และค่าเฉลี่ย (mean)

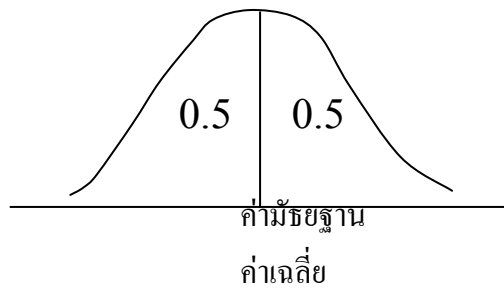
2.7.1 ค่าฐานนิยม

ค่าฐานนิยม คือ ค่าของข้อมูลที่มีความถี่มากที่สุดในการแจกแจงหนึ่ง เหมาะสำหรับข้อมูลที่แบ่งเป็นชั้น (categorical data) ซึ่งตัวแปรเป็นแบบแบ่งประเภท มีสเกลการวัดแบบแบ่งประเภท (nominal scale) เป็นค่าที่หายากที่สุดของการวัดแนวโน้มเข้าสู่ส่วนกลาง ถ้าวัดค่าฐานนิยมของข้อมูลที่เป็นแบบต่อเนื่อง (continuous data) โดยการจัดกลุ่มข้อมูลนั้นให้เป็นกลุ่ม (grouped data) แบ่งเป็นชั้น ๆ ที่มีความกว้างของชั้นไม่

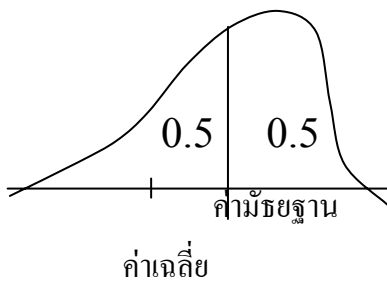
เท่ากัน อาจทำให้รูปทรงของการแจกแจงบิดเบือนไปจากที่เป็นจริงได้ ทำให้เห็นได้ชัดว่าต้องระมัดระวังในการตีความหมายข้อมูล

2.7.2 ค่ามัธยฐาน

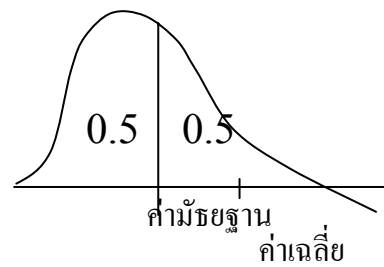
ค่ามัธยฐาน คือ ค่าของข้อมูลที่ตำแหน่งกลางของการแจกแจงที่มีจำนวนความถี่ของข้อมูลที่มีค่ามากกว่าหรือน้อยกว่าจุดนี้เป็นจำนวนความถี่เท่ากับครึ่งหนึ่งของจำนวนข้อมูลทั้งหมดเป็นจุดที่แบ่งการแจกแจงออกเป็น 2 ส่วน คือ ด้านซ้าย และด้านขวาเท่า ๆ กัน ค่ามัธยฐานเหมาะสมมากกับข้อมูลที่เป็นตัวเลขเชิงปริมาณที่มีการแจกแจงแบบเบ้ ตัวอย่างเช่น การแจกแจงแบบเบ้ซ้าย จะมีข้อมูลที่มีค่าน้อยเพียงเล็กน้อยเท่านั้นอยู่ทางทางด้านซ้ายของการแจกแจง แต่ข้อมูลที่มีค่าน้อยนี้มีผลอย่างมากในการคำนวณค่าเฉลี่ย คือทำให้ค่าเฉลี่ยต่ำลงหรือถูกดึงให้ไปทางด้านซ้ายที่ข้อมูลมีค่าน้อยนั้น และในทำนองเดียวกันการแจกแจงแบบเบ้ขวา จะทำให้ค่าเฉลี่ยสูงขึ้น หรือถูกดึงไปทางด้านขวาที่ข้อมูลมีค่ามากนั้น ดังนั้นค่าเฉลี่ยจึงไม่เหมาะสมที่จะเป็นค่ากลางของข้อมูล เพราะถูกดึงด้วยข้อมูลที่มีค่าน้อยหรือมากที่มีอยู่เพียงเล็กน้อยเท่านั้น จึงควรใช้ค่ามัธยฐานในการวัดค่ากลางของข้อมูลที่เป็นตัวเลขเชิงปริมาณที่มีการแจกแจงแบบเบ้



(ก) ตำแหน่งค่าเฉลี่ยและค่ามัธยฐานของการแจกแจงแบบสมมาตร



(ข) ตำแหน่งค่าเฉลี่ยและค่ามัธยฐานของการแจกแจงแบบเบ้ซ้าย



(ค) ตำแหน่งค่าเฉลี่ยและค่ามัธยฐานของการแจกแจงแบบเบ้ขวา

ภาพที่ 4.10 แสดงตำแหน่งของค่าเฉลี่ย และค่ามัธยฐานของการแจกแจง 3 แบบ

นอกเหนือจากค่ามัธยฐานที่บอกตำแหน่งกลางของข้อมูล ซึ่งแบ่งข้อมูลออกเป็น 2 ส่วนเท่า ๆ กัน เรายังมีเปอร์เซ็นต์ไทล์ (percentiles) และควอไทล์ (quartiles) ที่บอกตำแหน่ง ต่าง ๆ ของข้อมูลได้ด้วย การหาตำแหน่งของข้อมูลในการแจกแจงสำหรับข้อมูลที่เป็นตัวเลขเชิงปริมาณ เช่น คะแนนสอบวิชาหนึ่ง

เปอร์เซ็นต์ไทล์เป็นการแบ่งข้อมูลทั้งหมดออกเป็น 100 ส่วน ส่วนละเท่า ๆ กัน ดังนั้นตำแหน่งของข้อมูลที่เปอร์เซ็นต์ไทล์ที่ 50 เราก็เรียกว่าค่ามัธยฐานของการแจกแจงด้วยเหมือนกัน เพราะว่าเป็นจุดที่แบ่งการแจกแจงของข้อมูลออกเป็น 2 ส่วนเท่า ๆ กัน ตัวอย่างของเปอร์เซ็นต์ไทล์ จากการแจกแจงความถี่ของคะแนนสอบวิชาหลักการวางแผนการตลาดในตารางที่ 4.3 คะแนนสอบ 47 คะแนนคือตำแหน่งเปอร์เซ็นต์ไทล์ที่ 50.8 และคะแนนสอบ 45 คือตำแหน่งเปอร์เซ็นต์ไทล์ที่ 39.8 หมายความว่ามึนักเรียนที่สอบได้คะแนนน้อยกว่าหรือเท่ากับ 45 คะแนน จำนวนเท่ากับ 39.8% ของนักเรียนทั้งหมด

ควอไทล์เป็นการแบ่งข้อมูลออกเป็น 4 ส่วน ส่วนละเท่า ๆ กัน โดยที่ควอไทล์ที่ 1 (first quartile) แทนด้วย Q_1 เท่ากับเปอร์เซ็นต์ไทล์ที่ 25 ควอไทล์ที่ 2 (second quartile) แทนด้วย Q_2 เท่ากับเปอร์เซ็นต์ไทล์ที่ 50 และควอไทล์ที่ 3 (third quartile) แทนด้วย Q_3 เท่ากับเปอร์เซ็นต์ไทล์ที่ 75

ตัวอย่างการคิดควอไทล์จากข้อมูลคะแนนสอบในตารางที่ 4.3 มีนักเรียน 55 คน

(1) เรียงลำดับคะแนนจากน้อยไปมาก

(2) หาค่าลำดับที่ของคะแนนที่เปอร์เซ็นต์ไทล์ที่ 25 เท่ากับ $55 \times .25 = 13.8$ ซึ่งไม่เป็นเลขจำนวนเต็มดังนั้นจึงใช้เลขจำนวนเต็ม 14 เป็นลำดับของคะแนนของ Q_1

ดังนั้น $Q_1 = 41$ คะแนน

(3) หาค่าลำดับที่ของคะแนนที่เปอร์เซ็นต์ไทล์ที่ 50 เท่ากับ $55 \times .5 = 27.5$ ซึ่งไม่เป็นเลขจำนวนเต็มดังนั้นจึงใช้เลขจำนวนเต็ม 28 เป็นลำดับของคะแนนของ Q_2

ดังนั้น $Q_2 = 47$ คะแนน

(4) หาค่าลำดับที่ของคะแนนที่เปอร์เซ็นต์ไทล์ที่ 75 เท่ากับ $55 \times .75 = 41.3$ ซึ่งไม่เป็นเลขจำนวนเต็ม ดังนั้นจึงใช้เลขจำนวนเต็ม 42 เป็นลำดับของคะแนนของ Q_3

ดังนั้น $Q_3 = 52$ คะแนน

2.7.3 ค่าเฉลี่ย

ค่าเฉลี่ย คือ ค่ากลางของการแจกแจงของค่าของข้อมูล เป็นค่าที่ใช้มากที่สุดและมีประโยชน์มากของการวัดแนวโน้มเข้าสู่ส่วนกลาง เหมาะสำหรับข้อมูลแบบต่อเนื่องที่มีสเกลการวัดแบบช่วง (interval scale) และแบบอัตราส่วน (ratio scale) มีความหมายที่ซับซ้อนกว่าค่าฐานนิยมและค่ามัธยฐาน ค่าเฉลี่ยของตัวแปร X แทนด้วย \bar{x} คือ ผลบวกของค่าของข้อมูลทุกตัวคือ $\sum x_i$ หารด้วยจำนวนข้อมูลทั้งหมดคือ n มีสูตรการคำนวณคือ

$$\bar{x} = \frac{\sum x_i}{n} \quad ; \quad i = 1, 2, \dots, n$$

การทำความเข้าใจเกี่ยวกับความสัมพันธ์ของค่าฐานนิยม ค่ามัธยฐาน และค่าเฉลี่ยทำได้โดยอาศัยการพิจารณาจากการแจกแจงแบบปกติ คือ ถ้าข้อมูลมีการแจกแจงแบบปกติ การวัดแนวโน้มเข้าสู่ส่วนกลางจะได้ค่าฐานนิยม ค่ามัธยฐาน และค่าเฉลี่ยเท่ากัน หรืออยู่ที่ตำแหน่งเดียวกัน แต่ถ้าค่ามัธยฐานและค่าเฉลี่ยแตกต่างกัน แสดงว่าข้อมูลมีการแจกแจงแบบเบ้คือ ถ้าค่าเฉลี่ยมากกว่าค่ามัธยฐาน การแจกแจงจะเป็นแบบเบ้ขวา และถ้าค่าเฉลี่ยน้อยกว่าค่ามัธยฐาน การแจกแจงจะเป็นแบบเบ้ซ้าย เราสามารถคำนวณค่าความเบ้ (skewness) ของการแจกแจงได้จากสูตรคือ

$$\text{ความเบ้} = \frac{3(\text{ค่าเฉลี่ย} - \text{ค่ามัธยฐาน})}{\text{ค่าเบี่ยงเบนมาตรฐาน}}$$

2.8 การวัดการกระจาย (measures of dispersion)

นอกจากการอธิบายคุณลักษณะของประชากรด้วยการวัดแนวโน้มเข้าสู่ส่วนกลางแล้วเรายังต้องอธิบายคุณลักษณะเกี่ยวกับการแจกแจงของประชากรประกอบด้วย โดยการวัดการกระจายของข้อมูล ซึ่งมีอยู่หลายวิธี แต่ละวิธีจะเหมาะสมกับข้อมูลที่มีสเกลการวัดบางชนิดเท่านั้น จึงต้องเลือกใช้ค่าการวัดการกระจายให้เหมาะสมกับลักษณะของข้อมูล

2.8.1 ช่วงควอไทล์ (interquatile range)

เป็นค่าการกระจายของการแจกแจงโดยการคำนวณช่วงความห่างของข้อมูลที่บริเวณซึ่งมีความถี่มาก วิธีการนี้จะพิจารณาเฉพาะช่วงภายในข้อมูลที่อยู่ตรงกลางซึ่งคิดเป็น 50 เปอร์เซนต์ของข้อมูลทั้งหมด โดยที่ 25 เปอร์เซนต์ของข้อมูลทั้งหมดที่มีค่าต่ำสุดและ 25 เปอร์เซนต์ของข้อมูลทั้งหมดที่มีค่าสูงสุด ถูกตัดออกจากการพิจารณา เป็นการตัดข้อมูลส่วนริมของการแจกแจงออกไปจากการคำนวณเพราะอาจมีผลให้ค่าการกระจายของข้อมูลถูกบิดเบือนไป โดยทั่วไปใช้สำหรับวัดการกระจายของข้อมูลที่มีสเกลการวัดแบบอันดับ แต่อย่างไรก็ตามก็สามารถใช้ได้กับข้อมูลทั้งแบบต่อเนื่องและแบบไม่ต่อเนื่องด้วย มีสูตรการคำนวณคือ

$$\text{ช่วงควอไทล์} = Q_3 - Q_1$$

เมื่อ Q_1 และ Q_3 เรียกว่า ควอไทล์ที่ 1 และ 3 ตามลำดับ คำนวณได้จากการหาตำแหน่งที่ 25 เปอร์เซนต์ และ 75 เปอร์เซนต์ของการแจกแจงของข้อมูล ค่าของข้อมูลที่ตำแหน่ง 25 เปอร์เซนต์คือ ควอไทล์ที่ 1 และค่าของข้อมูลที่ตำแหน่ง 75 เปอร์เซนต์คือ ควอไทล์ที่ 3

2.8.2 พิสัย (range)

เป็นค่าการวัดการกระจายที่คำนวณง่ายที่สุดได้จากข้อมูลที่มีค่ามากที่สุด (max) ลบข้อมูลที่มีย่าน้อยที่สุด (min) ตัวอย่างเช่น คะแนนสอบวิชาหลักการวางแผนการตลาดของนักเรียน 55 คน มีคะแนนสูงสุด คือ 64 คะแนน และคะแนนต่ำสุดคือ 22 คะแนน ดังนั้นคำนวณหาค่าพิสัยได้ คือ

$$\text{พิสัย} = 64 - 22 = 42 \text{ คะแนน}$$

2.8.3 ค่าเบี่ยงเบนมาตรฐาน (standard deviation)

ใช้สำหรับวัดการกระจายของข้อมูลที่มีสเกลการวัดแบบช่วงและแบบอัตราส่วน มักแทนด้วย s หรือ sd มีสูตรการคำนวณคือ

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$$

เมื่อ x_i คือ ค่าของข้อมูลตัวที่ i

\bar{x} คือ ค่าเฉลี่ยของข้อมูลทั้งหมด

n คือ จำนวนข้อมูลทั้งหมด

ค่าเบี่ยงเบนมาตรฐานคำนวณได้จากผลบวกของความเบี่ยงเบนของข้อมูลแต่ละค่าจากค่าเฉลี่ย ซึ่งเท่ากับศูนย์ เราจึงยกกำลังสองเพื่อให้ค่าที่เป็นลบคูณกันได้เป็นบวก ทำให้ได้ผลบวกของค่าเบี่ยงเบนสัมบูรณ์ที่แสดงการกระจายของการแจกแจงของข้อมูล และเมื่อหารด้วยจำนวนข้อมูลทั้งหมดคือ n ก็จะได้ค่าเฉลี่ยของค่าเบี่ยงเบนสัมบูรณ์ทั้งหมด เป็นค่าการกระจายของการแจกแจงหนึ่ง เรียกว่า ความแปรปรวน (variance) มักแทนด้วย S^2 และเมื่อถอดรากที่สอง (square root) ของค่าความแปรปรวนก็จะได้ค่าเบี่ยงเบนมาตรฐาน

เราสามารถหาค่าเบี่ยงเบนมาตรฐานในการเปรียบเทียบระหว่างการแจกแจงแบบต่าง ๆ ได้ ถ้าค่าเบี่ยงเบนมาตรฐานมีค่าน้อย หมายความว่า การกระจายของการแจกแจงแคบ นั่นคือ ค่าของข้อมูลส่วนใหญ่อยู่รอบ ๆ ค่าเฉลี่ย และถ้ามีค่ามาก หมายความว่า การแจกแจงมีความกว้างมาก ส่วนใหญ่มักใช้กับข้อมูลที่มีการแจกแจงแบบปกติ

3. การเลือกใช้สถิติพรรณนาตามลักษณะของข้อมูล

3.1 ข้อมูลแบบแบ่งประเภท

- (1) การแจกแจงความถี่ ควรใช้กราฟแท่ง หรือกราฟวงกลม
- (2) การวัดแนวโน้มเข้าสู่ส่วนกลาง ควรใช้ค่าฐานนิยม
- (3) การกระจายของข้อมูล สามารถดูได้จากความถี่สัมพัทธ์ (relative frequency) ทำให้เห็นการกระจายของข้อมูลในแต่ละประเภทคิดเป็นสัดส่วนกับข้อมูลทั้งหมด มองเห็นภาพการกระจายของข้อมูลได้ชัดเจนขึ้น

3.2 ข้อมูลแบบอันดับ

- (1) การแจกแจงความถี่ ควรใช้กราฟแท่ง หรือกราฟวงกลม
- (2) การวัดแนวโน้มเข้าสู่ส่วนกลาง ควรใช้ค่ามัธยฐาน หรือค่าฐานนิยม
- (3) การวัดการกระจาย ควรใช้ช่วงควอไทล์

3.3 ข้อมูลแบบช่วง และแบบอัตราส่วน

- (1) การแจกแจงความถี่ ควรใช้กราฟฮิสโตแกรม หรือกราฟเส้น
- (2) การวัดแนวโน้มเข้าสู่ส่วนกลาง ควรใช้ค่าเฉลี่ยในกรณีที่มีข้อมูลมีการแจกแจงแบบสมมาตร และควรใช้ค่ามัธยฐานในกรณีที่มีข้อมูลมีการแจกแจงแบบเบ้
- (3) การวัดการกระจาย ควรใช้ค่าส่วนเบี่ยงเบนมาตรฐาน (standard deviation) ค่าสัมประสิทธิ์ความผันแปร (coefficient of variation range : C.V.)

4. คำสั่งที่ใช้สำหรับการวิเคราะห์ในสถิติพรรณนา (descriptives statistics)

มีคำสั่งที่สำคัญ 2 คำสั่ง คือ

- (1) คำสั่ง **Frequencies** ใช้สำหรับข้อมูลเชิงคุณภาพ และข้อมูลเชิงปริมาณ
- (2) คำสั่ง **Descriptives** ใช้สำหรับข้อมูลเชิงปริมาณ

4.1 การใช้คำสั่ง **Frequencies** สำหรับข้อมูลเชิงคุณภาพ

จากแฟ้มข้อมูล DataTest5.sav มีตัวแปร sex และ dep ที่มีสเกลการวัดแบบแบ่งประเภท คือ sex(1 = หญิง, 2 = ชาย) และ dep (1 = แผนกบุคคล , 2 = แผนกการเงิน , 3 = แผนกการตลาด) เราต้องการตารางแจกแจงความถี่ ร้อยละของตัวแปร sex และ dep และกราฟแท่ง มีขั้นตอนการใช้คำสั่ง ดังนี้

1. ไปที่เมนูบาร์ คลิกที่ **Analyze** , **Descriptive Statistics** , **Frequencies** จะได้หน้าต่าง **Frequencies**

2. ในหน้าต่าง **Frequencies** ในช่องซ้ายมือ คลิกที่ตัวแปร sex และ dep แล้วคลิกที่หัวลูกศร > หน้าช่อง **Variable(S)** : ตัวแปร sex และ dep จะย้ายไปอยู่ในช่อง **Variable(S)** :

เลือกคำสั่ง **Display Frequency Tables** เป็นคำสั่งให้แสดงตารางแจกแจงความถี่ ซึ่งโดยปกติโปรแกรมจะเลือกคำสั่งนี้อยู่แล้ว โดยจะมีเครื่องหมาย ในช่อง

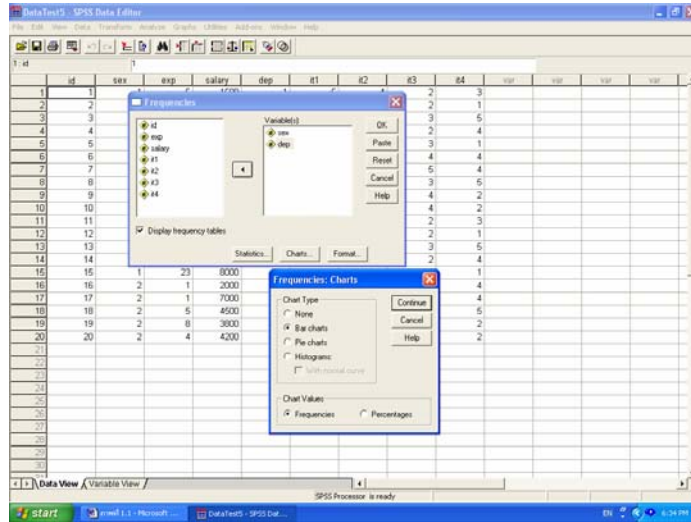
คลิกที่ปุ่ม **Charts...** จะได้หน้าต่าง **Frequencies : Charts** ดังภาพที่

4.10

3. ในหน้าต่าง **Frequencies : Charts** ภายในกรอบ **Chart Type** คลิกที่ **O Bar charts**

แล้วคลิกที่ปุ่ม Continue หน้าต่างนี้จะถูกปิดไป

4. ในหน้าต่าง Frequencies คลิกที่ปุ่ม OK จะได้ผลลัพธ์ดังแสดงในภาพที่ 4.11



ภาพที่ 4.10

Frequencies

Statistics

		sex	dep
N	Valid	20	20
	Missing	0	0

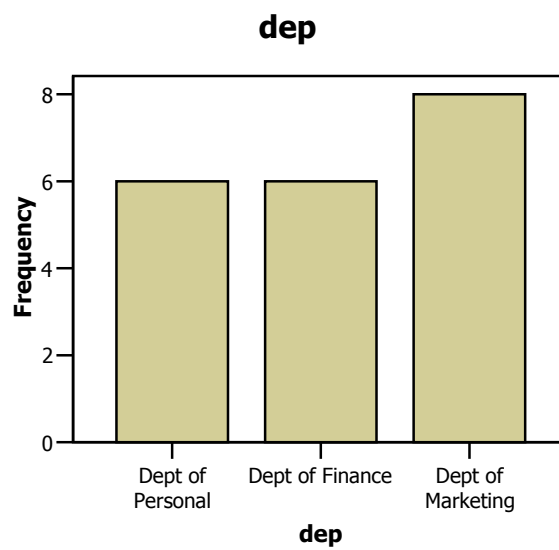
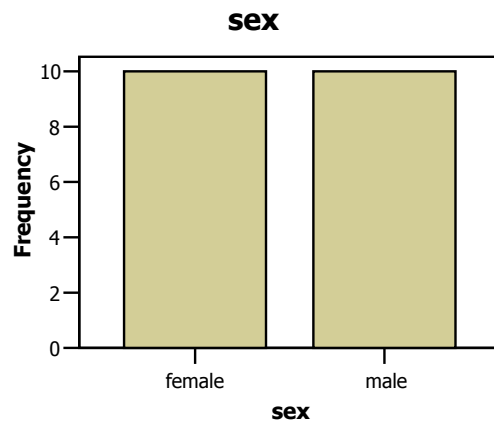
Frequency Table

sex

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	female	10	50.0	50.0	50.0
	male	10	50.0	50.0	100.0
Total		20	100.0	100.0	

dep

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Dept of Personal	6	30.0	30.0	30.0
	Dept of Finance	6	30.0	30.0	60.0
	Dept of Marketing	8	40.0	40.0	100.0
	Total	20	100.0	100.0	



ภาพที่ 4.11

ภาพที่ 4.11 แสดงผลลัพธ์ที่ได้จากการวิเคราะห์ในมิติพรรณนาคือ

1. ในตาราง **statistics** แสดงจำนวนข้อมูลที่ถูกต้องตามที่กำหนดไว้ (**N Valid**) ของตัวแปร **sex** (1 = female , 2 = male) และ **dep 1 = dep of personal** , **2 = dep of finance** , **3 = dep of marketing** เท่ากับ 20 และ 20 ตามลำดับ และจำนวนข้อมูลที่ผิดพลาด (**N Missing**) ของตัวแปร **sex** และ **dep** เท่ากับ 0 และ 0 ตามลำดับ

2. ในตาราง **sex** แสดงตารางแจกแจงความถี่ของข้อมูลที่ถูกต้องทั้งหมดของตัวแปร **sex** ที่มีลักษณะของข้อมูลเป็นชั้น (**categorical data**) และมีสเกลการวัดแบบแบ่งประเภท (**nominal scale**) คือ **female** และ **male** โดยที่ในตารางจะสรุปผลเป็นจำนวนความถี่ (**Frequency**) เปอร์เซ็นต์ (**Percent**) เปอร์เซ็นต์ที่ถูกต้อง (**Valid Percent**) และเปอร์เซ็นต์สะสม (**Cumulative Percent**)

และในตาราง **dep** แสดงตารางแจกแจงความถี่ของข้อมูลที่ถูกต้องทั้งหมดของตัวแปร **dep** ที่มีลักษณะของข้อมูลเป็นชั้น และมีสเกลการวัดแบบแบ่งประเภทคือ **dep of personal** , **dep of finance** , และ **dep of marketing** โดยที่ในตารางจะสรุปผลเช่นเดียวกับในตาราง **SEX**

3. กราฟแท่งของข้อมูลที่ถูกต้องของตัวแปร **sex** และ **dep**

4.2 การใช้คำสั่ง **Frequencies** สำหรับข้อมูลเชิงปริมาณ

ตัวอย่างเช่น จากแฟ้มข้อมูล **DataTest5.sav** มีตัวแปร **exp** และ **salary** ที่มีสเกลการวัดแบบอัตราส่วน (**ratio scale**) ซึ่งเป็นข้อมูลเชิงปริมาณ จึงไม่จำเป็นต้องแสดงตารางแจกแจงความถี่ของค่าทุกค่าของตัวแปร แต่อาจต้องการภาพการแจกแจงของตัวแปรด้วยกราฟฮิสโตแกรม ค่าเฉลี่ย มัชยฐาน ฐานนิยม ส่วนเบี่ยงเบนมาตรฐาน ความแปรปรวน ค่าน้อยที่สุด ค่ามากที่สุดของตัวแปร มีขั้นตอนการใช้คำสั่ง ดังนี้

1. ไปที่เมนูบาร์ คลิกที่ **Analyze** , **Descriptive Statistics** , **Frequencies ...** จะได้นหน้าต่าง **Frequencies**

2. ในหน้าต่าง **Frequencies** ในช่องซ้ายมือ คลิกที่ตัวแปร **exp** และ **salary** แล้วคลิกที่หัวลูกศร > หน้าช่อง **Variable(S)** : ตัวแปร **exp** และ **salary** จะย้ายไปอยู่ในช่องนี้

โดยปกติโปรแกรมจะเลือกคำสั่ง **Display frequency tables** ให้อยู่แล้ว แต่ในที่นี้เราไม่ต้องการตารางแจกแจงความถี่ของค่าทุกค่าของตัวแปร ดังนั้นจึงต้องคลิกใน เพื่อเอาเครื่องหมาย ✓ ออกไป

คลิกที่ปุ่ม **Statistics...** จะได้นหน้าต่าง **Frequencies : Statistics** ดังภาพที่ 4.12

3. ในหน้าต่าง **Frequencies : Statistics** ภายในกรอบ **Central Tendency** คลิกที่คำสั่ง **Mean** , **Median** , **Mode**

ภายในกรอบ **Dispersion** คลิกที่คำสั่ง **Std. Deviation** , **Variance** , **Minimum** , **Maximum**
แล้วคลิกที่ปุ่ม **Continue** หน้าต่างนี้จะถูกปิดไป

4. ในหน้าต่าง **Frequencies** คลิกที่ปุ่ม **Charts....** จะได้นหน้าต่าง **Frequencies : Charts**

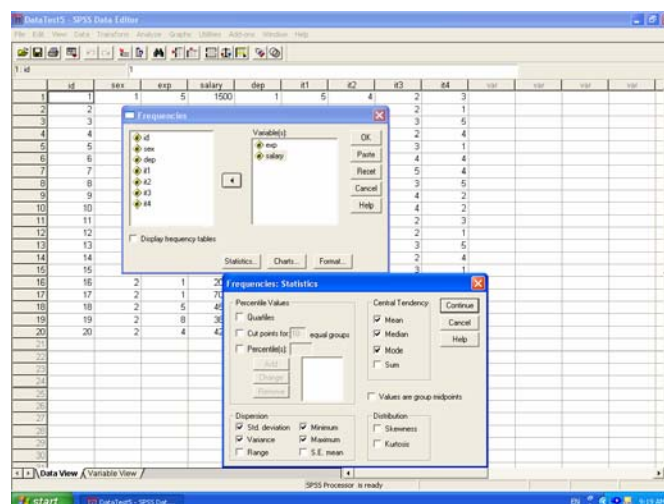
5. ในหน้าต่าง **Frequencies : Charts**

ภายในกรอบ **Chart Type** คลิกที่ **Histograms**

แล้วคลิกที่ปุ่ม **Continue** หน้าต่างนี้จะถูกปิดไป

6. ในหน้าต่าง **Frequencies**

คลิกที่ปุ่ม **OK** จะได้ผลลัพธ์ดังแสดงในภาพที่ 4.13



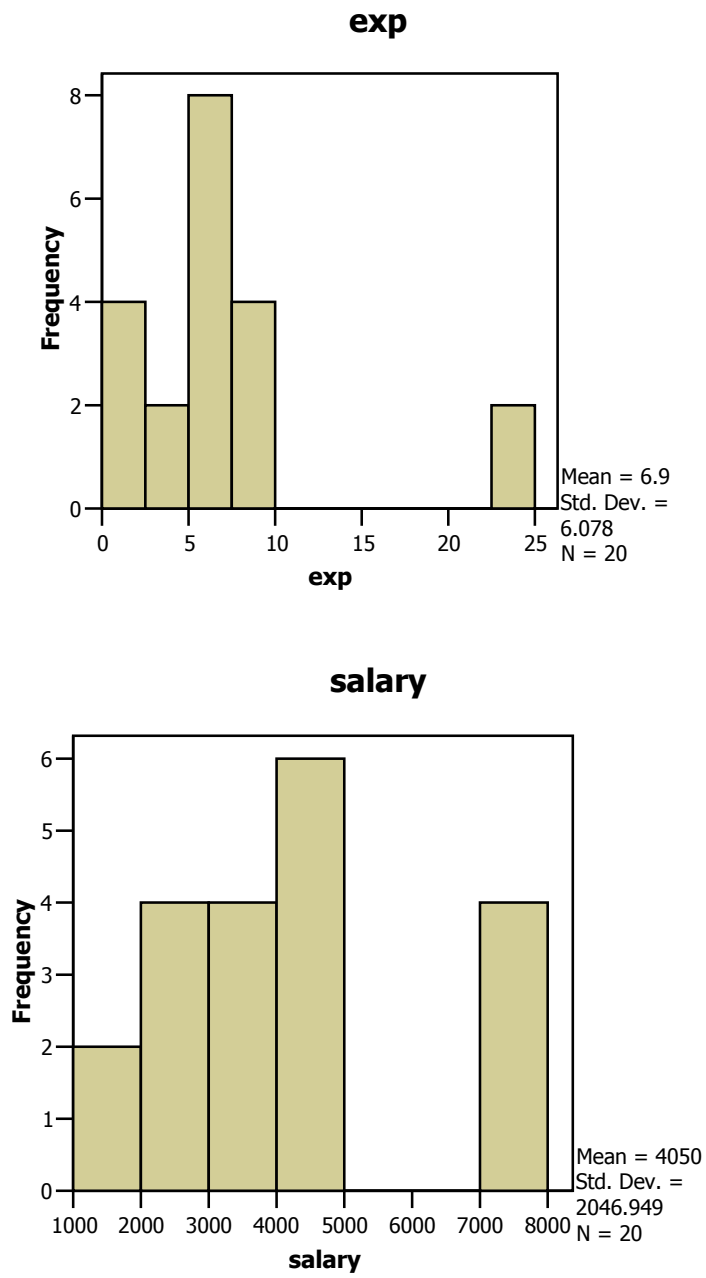
ภาพที่ 4.12

Frequencies**Statistics**

		exp	salary
N	Valid	20	20
	Missing	0	0
Mean		6.90	4050.00
Median		5.50	3900.00
Mode		1 ^a	2000
Std. Deviation		6.078	2046.949
Variance		36.937	4190000
Minimum		1	1500
Maximum		23	8000

a. Multiple modes exist. The smallest value is shown

Histogram



ภาพที่ 4.13

ภาพที่ 4.13 แสดงผลลัพธ์ที่ได้จากการวิเคราะห์ในสถิติพรรณนาคือ

1. ในตาราง **Statistics** แสดงผลสรุปรวมของข้อมูลสำหรับตัวแปร **exp** และ **salary** เป็นค่าต่าง ๆ คือ

- จำนวนข้อมูลที่ต้องตามที่กำหนดไว้ (**N Valid**)
- จำนวนข้อมูลที่ผิดพลาด (**N Missing**)
- ค่าเฉลี่ย (**Mean**)
- ค่ามัธยฐาน (**Median**)
- ค่าฐานนิยม (**Mode**)
- ค่าเบี่ยงเบนมาตรฐาน (**Std. Deviation**)
- ค่าความแปรปรวน (**Variance**)
- ค่าต่ำสุด (**Minimum**)
- ค่าสูงสุด (**Maximum**)

2. กราฟแท่งของข้อมูลที่ต้องของตัวแปร **exp** และ **salary**

4.3 การใช้คำสั่ง **Descriptives** สำหรับข้อมูลเชิงปริมาณ

ตัวอย่างเช่น จากเพิ่มข้อมูล **DataTest5.sav** มีตัวแปร **exp** และ **salary** ที่มีสเกลการวัดแบบอัตราส่วน ซึ่งเป็นข้อมูลเชิงปริมาณ ผลลัพธ์ที่ได้จากการใช้คำสั่ง **Descriptives** จะให้ค่าสถิติต่าง ๆ คล้ายกับที่ได้จากการใช้คำสั่ง **Frequencies** แต่ไม่มีค่าฐานนิยม มัธยฐาน ควอไทล์ เปอร์เซ็นต์ไทล์ และไม่มีตารางแจกแจงความถี่ ซึ่งมีอยู่ในคำสั่งของ **Frequencies** เท่านั้น เราต้องการหาค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน ความแปรปรวน และสัมประสิทธิ์ความเบ้ของการแจกแจงข้อมูล มีขั้นตอนการใช้คำสั่งดังนี้

1. ไปที่เมนูบาร์ คลิกที่ **Analyze** , **Descriptive Statistics** , **Descriptives...** จะได้นหน้าต่าง **Descriptives**

2. ในหน้าต่าง **Descriptives** ในช่องซ้ายมือ คลิกที่ตัวแปร **exp** และ **salary** แล้วคลิกที่หัวลูกศร > หน้าช่อง **Variable(S)** : ตัวแปร **exp** และ **salary** จะย้ายไปอยู่ในช่องนี้

คลิกที่ปุ่ม **Options...** จะได้นหน้าต่าง **Descriptives : Options**

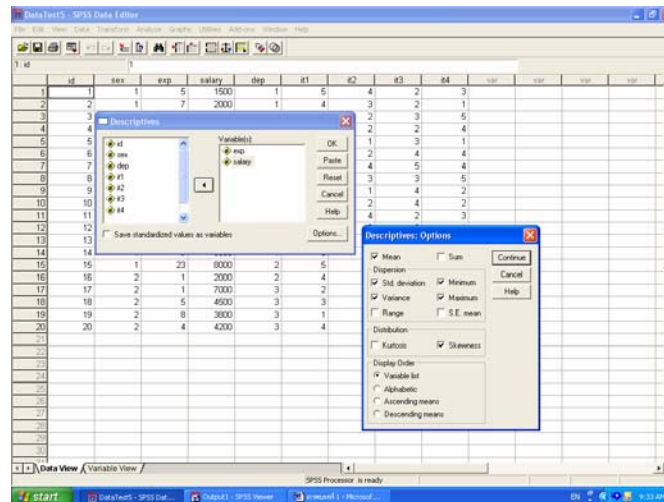
3. ในหน้าต่าง **Descriptives : Options** โปรแกรมได้กำหนดค่าสถิติต่าง ๆ ไว้ให้แล้วคือ ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน ค่าต่ำสุด ค่าสูงสุด ดังนั้น ค่าสถิติที่เราต้องการ

นอกเหนือจากนี้คือ ความแปรปรวน และสัมประสิทธิ์ความเบ้ (Skewness) ของการแจกแจงข้อมูล จึงต้องคลิกที่ Variance และ Skewness ดังภาพที่ 4.14

แล้วคลิกปุ่ม Continue หน้าต่างนี้จะถูกปิดไป

4. ในหน้าต่าง Descriptives

คลิกที่ปุ่ม OK จะได้ผลลัพธ์ดังภาพที่ 4.15



ภาพที่ 4.14

Descriptives

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std.	Variance	Skewness	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error
exp	20	1	23	6.90	6.078	36.937	1.946	.512
salary	20	1500	8000	4050.00	2046.949	4190000	.728	.512
Valid N (listwise)	20							

ภาพที่ 4.15

ภาพที่ 4.15 ในตาราง Descriptive Statistics แสดงผลสรุปรวมของข้อมูลสำหรับตัวแปร **exp** และ **salary** เป็นค่าสถิติต่าง ๆ คือ จำนวนข้อมูล (N) ค่าน้อยที่สุด ค่ามากที่สุด ค่าเฉลี่ย ค่าเบี่ยงเบนมาตรฐาน ค่าความแปรปรวน ค่าความเบ้ โดยบอกค่าความคลาดเคลื่อนมาตรฐาน (Std. Error) ของความเบ้ด้วย

