การวิเคราะห์ข้อมูลเบื้องต้น

ผศ.ดร.ธนวัฒน์ ศรีศิริวัฒน์

สาขาวิชาคณิตศาสตร์

คณะครุศาสตร์

มหาวิทยาลัยราชภัฏสวนสุนันทา

การแจกแจงความถี่

การแจกแจงความถี่ (frequency distribution) เป็นวิธีการทางสถิติที่ใช้ในการจัดข้อมูลที่เก็บรวบรวมมาให้อยู่เป็นกลุ่มๆ หรือช่วง ซึ่งในแต่ละกลุ่มถูกแยกออกจากกันอย่างเป็นหมวดหมู่ มองเห็นได้ง่าย เหมาะอย่างยิ่งสำหรับข้อมูลที่มีจำนวนมากๆ ปกติแล้วการแจกแจงความถี่นิยมทำเป็นรูปตาราง จึงเรียกว่า “**ตารางแจกแจงความถี่**” ซึ่งมีขั้นตอนในการสร้าง ดังนี้

การแจกแจงความถี่

ขั้นที่ 1 หาพิสัย (range) เป็นค่าความแตกต่างระหว่างค่าสูงสุดและค่าต่ำสุดของข้อมูล โดยคำนวณหาได้จาก

$$R = X_{max} - X_{min}$$

เมื่อ	R	แทน	พิสัย
	X_{max}	แทน	ค่าของข้อมูลที่มีค่าสูงสุด
	X_{min}	แทน	ค่าของข้อมูลที่มีค่าต่ำสุด

การแจกแจงความถี่

ขั้นที่ 2 กำหนดจำนวนชั้นหรือกลุ่ม หรืออาจคำนวณหาจำนวนชั้นจากสูตรโดย Sturges’rule ดังนี้

$$k = 1 + 3.322\left(\log_{10} n\right)$$

เมื่อ	k	แทน	จำนวนชั้น
	n	แทน	จำนวนข้อมูลทั้งหมดที่มีอยู่

การแจกแจงความถี่

ขั้นที่ 3 หาความกว้างของชั้นหรือ “อันตรภาคชั้น” (class interval) ซึ่งหาได้จาก

$$W = \frac{R}{k}$$

เมื่อ W แทน ความกว้างของแต่ละชั้น

ขั้นที่ 4 ดำเนินการขีดรอยคะแนน (tally) ของข้อมูลทุกรายการ เพื่อหาความถี่ของข้อมูล ในแต่ละชั้น

ตัวอย่าง อายุของคน 40 คน ที่เข้าร่วมประชุมสัมมนาทางวิชาการครั้งหนึ่ง เป็นดังนี้

20	52	80	24	36	84	35	28	24	42
45	31	81	24	33	47	80	83	55	56
58	32	24	60	67	55	90	59	71	78
54	24	81	93	41	52	82	24	98	47

จงสร้างตารางแจกแจงความถี่ของอายุของผู้เข้าร่วมประชุมสัมมนาทางวิชาการ8*

วิธีทำ จากข้อมูลข้างต้น ค่าสูงสุด คือ 98 ค่าต่ำสุด คือ 20

ดังนั้น พิสัย คือ $R = 98 - 20 = 78$

จำนวนชั้นหาได้จาก $k = 1 + 3.322(\log_{10} n)$

จะได้ $k = 1 + 3.322(\log_{10} 40)$

$k \approx 6$

และหาขนาดของความกว้างของชั้น คือ $W = \frac{R}{k}$

จะได้ $W = \frac{78}{6} = 13$

ดังนั้น ความกว้างของชั้น เป็น 13 และได้ช่วงของข้อมูลเป็น

20 – 32, 33 – 45, 46 – 58, 59 – 71, 72 – 84 และ 85 – 98

หลังจากนั้นดำเนินการสร้างตารางและขีดรอยคะแนนแล้วนับจำนวนข้อมูลที่มีค่าตกอยู่ในแต่ละชั้น เพื่อหาความถี่และจะได้ตารางการแจกแจงความถี่ของข้อมูลอายุของผู้เข้าร่วม ประชุมสัมมนาทางวิชาการดังนี้

อายุ (ปี)	**รอยขีด**	**ความถี่**
20 – 32	~~////~~ ~~////~~	10
33 – 45	~~////~~ /	6
46 – 58	~~////~~ ////	9
59 – 71	////	4
72 – 84	~~////~~ ///	8
85 – 98	///	3
รวม		40

นอกจากการแจกแจงความถี่แล้วผู้ศึกษาอาจนำเสนอโดยการแจกแจงความถี่สะสมการแจกแจงความถี่สัมพัทธ์ และการแจกแจงความถี่สะสมสัมพัทธ์ โดยมีรายละเอียดดังนี้

ความถี่สะสม (cumulative frequency) เป็นผลรวมของความถี่ของค่านั้นหรืออันตรภาคชั้นนั้นกับอันตรภาคชั้นที่มีช่วงคะแนนต่ำกว่าทั้งหมด

ความถี่สัมพัทธ์ (relative frequency) เป็นอัตราส่วนของความถี่ของแต่ละชั้นข้อมูลกับความถี่รวม หรือจำนวนข้อมูลทั้งหมด

ความถี่สะสมสัมพัทธ์ (relative cumulative frequency) เป็นผลรวมสะสมของความถี่สัมพัทธ์สะสมจากชั้นที่มีค่าของข้อมูลต่ำสุดไปยังค่าของข้อมูลสูงสุด

ตารางแจกแจงความถี่ ความถี่สะสม ความถี่สัมพัทธ์ และความถี่สะสมสัมพัทธ์

จำแนกตามช่วงอายุของผู้เข้าร่วมสัมมนาทางวิชาการ

อายุ (ปี)	ความถี่ (f)	ความถี่สะสม (F)	ความถี่สัมพัทธ์	ความถี่สะสมสัมพัทธ์
20 – 32	10	10	0.250	0.250
33 – 45	6	16	0.150	0.400
46 – 58	9	25	0.225	0.625
59 – 71	4	29	0.100	0.725
72 – 84	8	37	0.200	0.925
85 – 98	3	40	0.075	1.000
รวม	40		1.000	

นอกจากการแจกแจงความถี่ข้างต้นแล้วยังสามารถแจกแจงความถี่โดยใช้กราฟดังนี้

1) ฮิสโทแกรม (histogram) เป็นแผนภูมิที่มีลักษณะเป็นรูปสี่เหลี่ยมมุมฉากวางเรียงติดกันบนแกนนอน โดยมีแกนนอนแทนค่าของตัวแปร มีความกว้างขนาดเท่ากับความกว้างของอันตรภาคชั้น ความสูงแสดงความถี่หรือความถี่สัมพัทธ์ และในบางครั้งค่าสังเกตไม่เป็นค่าต่อเนื่อง ดังนั้นในการสร้างฮิสโทแกรมเพื่อให้รูปสี่เหลี่ยมมุมฉากวางเรียงติดกัน จึงต้องหาขอบล่างและขอบบนของแต่ละอันตรภาคชั้นก่อน

2) รูปหลายเหลี่ยมของความถี่ (the frequency polygon) เป็นรูปที่เกิดจากการลากเส้นต่อจุดกึ่งกลางของแต่ละชั้น ด้านบนของแท่งฮิสโทแกรม พื้นที่ทั้งหมดภายใต้รูปหลายเหลี่ยมความถี่จะเท่ากับพื้นที่ทั้งหมดภายใต้ฮิสโทแกรม

การวัดแนวโน้มเข้าสู่ส่วนกลาง

1. ค่าเฉลี่ยเลขคณิต (arithmetic mean) อาจเรียกสั้นๆ ว่าค่าเฉลี่ย (mean) เป็นการหาค่ากลางของข้อมูลชนิดหนึ่ง ซึ่งเป็นค่าที่ได้จากการเฉลี่ยข้อมูลทั้งหมด

1.1 ค่าเฉลี่ยเลขคณิตกรณีข้อมูลไม่แจกแจงความถี่ หาได้โดยการหารผลรวมของข้อมูลทั้งหมดด้วยจำนวนข้อมูล นั่นคือ ถ้าให้ เป็นข้อมูลขนาด จากประชากร และเป็นข้อมูลขนาด จากกลุ่มตัวอย่าง จะได้

ค่าเฉลี่ยเลขคณิตของประชากร คือ $\mu = \frac{\sum_{i=1}^{N} X_i}{N}$

ค่าเฉลี่ยเลขคณิตของกลุ่มตัวอย่าง คือ $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$

ตัวอย่าง จากการสุ่มคะแนนสอบวิชาคณิตศาสตร์ครั้งหนึ่งของนักเรียน 10 คน ได้ข้อมูลดังนี้ 10, 15, 15, 12, 10, 8, 18, 9, 10, 14 คะแนน จงหาคะแนนเฉลี่ยของนักเรียนในการสอบครั้งนี้

วิธีทำ ด้วยคะแนนสอบเป็นคะแนนของนักเรียนที่สุ่มมาจำนวน 10 คน จะได้

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

$$\bar{X} = \frac{[10+15+15+12+10+8+18+9+10+14]}{10}$$

$$\bar{X} = 12.10$$

ดังนั้น คะแนนเฉลี่ยในการสอบเท่ากับ 12.10 คะแนน

สมบัติที่สำคัญของค่าเฉลี่ยเลขคณิต

ในการศึกษาค่าเฉลี่ยเลขคณิตจะพบว่ามีสมบัติที่สำคัญดังนี้

1) $\sum\left(X_i - \bar{X}\right) = 0$

2) $\bar{X}_{(X \pm k)} = \bar{X}_X \pm k$

3) $\bar{X}_{(kX)} = k\bar{X}_X$

4) $\bar{X}_{\left(\frac{x}{k}\right)} = \left(\frac{1}{k}\right)\bar{X}_x$

5) ค่าเฉลี่ยเลขคณิตเป็นค่าสถิติที่ใช้กับข้อมูลในระดับอันตรภาคและอัตราส่วน

6) ค่าเฉลี่ยเลขคณิตไม่เหมาะกับชุดข้อมูลที่มีค่าสูงมากๆ หรือต่ำมากๆ

เนื่องจากจะทำให้ค่าเฉลี่ยเลขคณิตที่คำนวณได้ไม่เป็นตัวแทนที่ดีของข้อมูลนั้น

สมบัติที่สำคัญของค่าเฉลี่ยเลขคณิต

7) กรณีที่ค่าสังเกตหรือข้อมูลแต่ละตัวมีน้ำหนักความสำคัญไม่เท่ากัน เช่น การคำนวณหาผลการเรียนเฉลี่ย ซึ่งในแต่ละวิชามีน้ำหนักหรือจำนวนหน่วยกิตไม่เท่ากัน ถ้าใช้ค่าเฉลี่ยเลขคณิตอาจทำให้เกิดความคลาดเคลื่อนได้ จึงควรใช้ “ค่าเฉลี่ยเลขคณิตถ่วงน้ำหนัก” (weighted arithmetic mean) จะทำให้ได้ค่ากลางที่เป็นตัวแทนของข้อมูลที่ดี

$$\overline{X}_w = \frac{\sum_{i=1}^{n} w_i X_i}{\sum_{i=1}^{n} w_i}$$

เมื่อ $\overline{X}_w$ แทน ค่าเฉลี่ยเลขคณิตถ่วงน้ำหนัก

w_i แทน น้ำหนักความสำคัญของ x_i

สมบัติที่สำคัญของค่าเฉลี่ยเลขคณิต

8) กรณีที่มีข้อมูลหลายชุดแต่ละชุดก็มีค่าเฉลี่ยเลขคณิต ถ้าต้องการทราบค่าเฉลี่ยเลขคณิตของข้อมูลทั้งหมด สามารถหาได้จาก “ค่าเฉลี่ยเลขคณิตรวม” (combined arithmetic mean) มีสูตรดังนี้

$$\bar{\bar{X}} = \frac{\sum_{i=1}^{k} n_i \bar{X}_i}{\sum_{i=1}^{k} n_i}$$

เมื่อ $\bar{\bar{X}}$ แทน ค่าเฉลี่ยเลขคณิตรวม

$\bar{X}_i$ แทน ค่าเฉลี่ยเลขคณิตของข้อมูลชุดที่ i

การวัดแนวโน้มเข้าสู่ส่วนกลาง

1.2 ค่าเฉลี่ยเลขคณิตกรณีข้อมูลแจกแจงความถี่ เป็นการหาค่าเฉลี่ยเลขคณิตในกรณีที่มีข้อมูลจำนวนมากและแจกแจงความถี่ คือถ้ากำหนดให้ f_1 เป็นความถี่ของข้อมูลในอันตรภาคชั้นที่ 1 ที่มี x_1 เป็นจุดกึ่งกลางชั้น f_2 และเป็นความถี่ของข้อมูลในอันตรภาคชั้นที่ 2 ที่มี x_2 เป็นจุดกึ่งกลางชั้นไปเรื่อยๆ ถึง f_k เป็นความถี่ของข้อมูลในอันตรภาคชั้นที่ k ที่มี x_k เป็นจุดกึ่งกลางชั้น จะได้

การวัดแนวโน้มเข้าสู่ส่วนกลาง

ค่าเฉลี่ยเลขคณิตของประชากรคือ

$$\mu = \frac{\sum_{i=1}^{k} f_i X_i}{N}$$

ค่าเฉลี่ยเลขคณิตของกลุ่มตัวอย่างคือ

$$\bar{X} = \frac{\sum_{i=1}^{k} f_i X_i}{n}$$

การวัดแนวโน้มเข้าสู่ส่วนกลาง

ในกรณีที่ข้อมูลมีค่ามากๆ อาจใช้เทคนิคการลดคะแนน โดยใช้สูตรการคำนวณ ดังนี้

$$\bar{X} = a + i\left[\frac{\sum_{i=1}^{k} f_i d_i}{n}\right]$$

เมื่อ a แทน คะแนนเฉลี่ยที่สมมติขึ้น

d_i แทน ค่าของจุดกึ่งกลางชั้นลบด้วยคะแนนเฉลี่ยสมมติ หารด้วยความกว้างของอันตรภาคชั้น

i แทน ความกว้างของอันตรภาคชั้น

การวัดแนวโน้มเข้าสู่ส่วนกลาง

2. มัธยฐาน (median : Med) เป็นค่าของข้อมูลที่อยู่กึ่งกลางของข้อมูลทั้งชุดเมื่อเรียงลำดับข้อมูลชุดนั้นจากน้อยไปมากหรือจากมากไปน้อย

2.1 มัธยฐานกรณีข้อมูลไม่แจกแจงความถี่ หาได้โดยการเรียงลำดับข้อมูลจากน้อยไปมากหรือจากมากไปน้อย แล้วพิจารณาว่า ถ้าจำนวนข้อมูลเป็นจำนวนคี่ มัธยฐานคือค่าที่อยู่ในตำแหน่งกึ่งกลางของข้อมูลทั้งหมด แต่ถ้าจำนวนข้อมูลเป็นจำนวนคู่ มัธยฐานคือค่าที่อยู่ในตำแหน่งกึ่งกลางของข้อมูลทั้งหมดบวกกันแล้วหารด้วย 2 และถ้าข้อมูลมีจำนวนมากสามารถหาตำแหน่งของมัธยฐาน โดยคำนวณจากสูตร ตำแหน่งของ $\text{Med} = \frac{(N+1)}{2}$

การวัดแนวโน้มเข้าสู่ส่วนกลาง

2.2 มัธยฐานกรณีข้อมูลแจกแจงความถี่ สามารถหาได้โดยสร้างตารางแจกแจงความถี่สะสม แล้วพิจารณาชั้นที่มีตำแหน่งของ Med คือข้อมูลตัวที่ $\frac{N}{2}$ และคำนวณหามัธยฐานจากสูตร

$$\text{Med} = L + i\left[\frac{\frac{N}{2} - F_L}{f_M}\right]$$

เมื่อ	L	แทน	ขอบล่างของชั้นที่มีมัธยฐาน
	N	แทน	จำนวนข้อมูลทั้งหมด
	F_L	แทน	ความถี่สะสมของชั้นที่มีค่าต่ำกว่าชั้นที่มีมัธยฐาน
	f_M	แทน	ความถี่ของชั้นที่มีมัธยฐาน
	i	แทน	ความกว้างของอันตรภาคชั้น

การวัดแนวโน้มเข้าสู่ส่วนกลาง

3. ฐานนิยม (mode:Mo) คือค่าของข้อมูลที่มีความถี่สูงสุด หรือเกิดบ่อยครั้งที่สุด ข้อมูล บางชุด อาจจะไม่มีฐานนิยมเลยก็ได้ ในกรณีที่มีจำนวนข้อมูลมากอาจมีฐานนิยมมากกว่า 1 ค่าได้

3.1 ฐานนิยมกรณีข้อมูลไม่แจกแจงความถี่ สามารถหาฐานนิยมโดยพิจารณาค่าหรือข้อมูลที่มีความถี่สูงสุด

การวัดแนวโน้มเข้าสู่ส่วนกลาง

3.2 ฐานนิยมกรณีข้อมูลแจกแจงความถี่ สามารถหาได้จากการคำนวณ โดยใช้สูตรในการคำนวณดังนี้

$$Mo = L + i\left(\frac{d_1}{d_1 + d_2}\right)$$

เมื่อ d_1 แทน ผลต่างของความถี่ของชั้นที่ฐานนิยมอยู่กับความถี่ของชั้นต่ำกว่า

d_2 แทน ผลต่างของความถี่ของชั้นที่ฐานนิยมอยู่กับความถี่ของชั้นสูงกว่า

L แทน ขอบล่างของชั้นที่มีความถี่สูงสุด

i แทน ความกว้างของอันตรภาคชั้น

ตารางที่ 3.3 การเลือกใช้การวัดแนวโน้มเข้าสู่ส่วนกลาง

การวัดแนวโน้มเข้าสู่ส่วนกลาง	การเลือกใช้
ค่าเฉลี่ยเลขคณิต	- เป็นค่าสถิติที่ใช้กับข้อมูลในระดับอันตรภาค และระดับ อัตราส่วน - เป็นค่าที่ได้จากการเฉลี่ยข้อมูลทั้งหมด - ไม่เหมาะกับชุดข้อมูลที่มีค่าสูงผิดปกติ หรือต่ำผิดปกติ เนื่องจากจะทำให้ค่าเฉลี่ยเลขคณิตที่คำนวณได้ไม่เป็นตัว แทนที่ดีของข้อมูล
มัธยฐาน	- เป็นค่าสถิติที่ใช้กับข้อมูลในระดับเรียงอันดับระดับอันตรภาค และระดับอัตราส่วน - สามารถใช้กับชุดข้อมูลที่มีค่าสูงผิดปกติหรือต่ำผิดปกติเนื่องจากไม่กระทบต่อการคำนวณ
ฐานนิยม	- เป็นค่าสถิติที่ใช้กับข้อมูลทุกระดับ - สามารถใช้กับชุดของข้อมูลที่มีค่าสูงผิดปกติ หรือต่ำผิดปกติ เนื่องจากไม่กระทบต่อการคำนวณ - เหมาะสำหรับใช้ข้อมูลที่มีการซ้ำกันมากๆ

การวัดการกระจาย

การวัดการกระจาย (measure of variation) เป็นการพิจารณาถึงลักษณะการกระจายของข้อมูลว่ามีการกระจายมากน้อยเพียงใดเพื่อประกอบการเลือกข้อมูลเมื่อข้อมูลแต่ละชุดมีการวัดแนวโน้มเข้าสู่ส่วนกลางเท่ากัน โดยแบ่งได้ 2 วิธี คือ การวัดการกระจายสัมบูรณ์ และการวัดการกระจายสัมพัทธ์ มีรายละเอียดดังนี้

การวัดการกระจาย

1. การวัดการกระจายสัมบูรณ์ (measure of absolute variation) เป็นการวัดการกระจายของข้อมูลเพียงชุดเดียวเพื่อเปรียบเทียบความแตกต่างของค่าสังเกตของข้อมูล โดยการวัดการกระจาย ที่นิยมใช้มี 5 ชนิดดังนี้

1.1 พิสัย (range) เป็นค่าความแตกต่างระหว่างค่าสูงสุดและค่าต่ำสุดของข้อมูลหนึ่งๆ ในกรณีไม่แจกแจงความถี่ หาได้จาก

พิสัย = ค่าสูงสุด – ค่าต่ำสุด

$$(R = X_{max} - X_{min})$$

การวัดการกระจาย

1.2 ส่วนเบี่ยงเบนควอร์ไทล์ (quartile deviation) เป็นค่าครึ่งหนึ่งของผลต่างระหว่างควอร์ไทล์ที่ 3 กับควอร์ไทล์ที่ 1 โดยการหาส่วนเบี่ยงเบนควอร์ไทล์ทั้งกรณีที่ยังไม่แจกแจงความถี่ และกรณีที่การแจกแจงความถี่ หาได้จาก

$$Q.D. = \frac{Q_3 - Q_1}{2}$$

เมื่อ $Q.D.$ แทน ส่วนเบี่ยงเบนควอร์ไทล์

Q_3 แทน ค่าของข้อมูล ณ ตำแหน่งควอร์ไทล์ที่ 3

Q_1 แทน ค่าของข้อมูล ณ ตำแหน่งควอร์ไทล์ที่ 1

การวัดการกระจาย

1.2 ส่วนเบี่ยงเบนควอร์ไทล์ (quartile deviation) เป็นค่าครึ่งหนึ่งของผลต่างระหว่างควอร์ไทล์ที่ 3 กับควอร์ไทล์ที่ 1 โดยการหาส่วนเบี่ยงเบนควอร์ไทล์ทั้งกรณีที่ยังไม่แจกแจงความถี่ และกรณีที่การแจกแจงความถี่ หาได้จาก

$$Q.D. = \frac{Q_3 - Q_1}{2}$$

เมื่อ Q.D. แทน ส่วนเบี่ยงเบนควอร์ไทล์

Q_3 แทน ค่าของข้อมูล ณ ตำแหน่งควอร์ไทล์ที่ 3

Q_1 แทน ค่าของข้อมูล ณ ตำแหน่งควอร์ไทล์ที่ 1

การวัดการกระจาย

1.3 ส่วนเบี่ยงเบนเฉลี่ย (mean deviation หรือ average deviation) เป็นค่าที่ได้จากการเฉลี่ยของค่าเบี่ยงเบนของข้อมูลแต่ละตัวที่เบี่ยงเบนไปจากค่าเฉลี่ยของข้อมูลชุดนั้น ซึ่งแตกต่างกับพิสัย และส่วนเบี่ยงเบนควอร์ไทล์ ที่คำนวณเฉพาะค่าของข้อมูลบางค่าเท่านั้นโดยแบ่งได้ 2 กรณีดังนี้

การวัดการกระจาย

1.3.1 ส่วนเบี่ยงเบนเฉลี่ยในกรณีข้อมูลไม่แจกแจงความถี่ สามารถคำนวณได้จากสูตร

$$M.D. = \frac{\sum_{i=1}^{k} \left| X_i - \bar{X} \right|}{n}$$

เมื่อ	MD.	แทน	ส่วนเบี่ยงเบนเฉลี่ย
	X_i	แทน	ค่าหรือคะแนนของข้อมูลที่ i
	$\bar{X}$	แทน	ค่าเฉลี่ยเลขคณิตของข้อมูลชุดนั้น
	n	แทน	จำนวนข้อมูลทั้งหมด

การวัดการกระจาย

1.3.2 ส่วนเบี่ยงเบนเฉลี่ยในกรณีข้อมูลแจกแจงความถี่

สามารถคำนวณได้จากสูตร

$$M.D. = \frac{\sum_{i=1}^{k} f_i \left| X_i - \bar{X} \right|}{n}$$

การวัดการกระจาย

1.4 ส่วนเบี่ยงเบนมาตรฐาน (standard deviation) เป็นการวัดการกระจายของข้อมูลที่มักนิยมใช้ ด้วยเป็นการหารากที่สองของค่าเฉลี่ยของกำลังสองของค่าเบี่ยงเบนของข้อมูลแต่ละตัวที่เบี่ยงเบนจากค่าเฉลี่ยของข้อมูลชุดนั้น โดยแบ่งได้ 2 กรณี มีรายละเอียดดังนี้

การวัดการกระจาย

1.4.1 ส่วนเบี่ยงเบนมาตรฐานกรณีข้อมูลไม่แจกแจงความถี่

สามารถคำนวณได้จากสูตรดังนี้

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N}(x_i-\mu)^2}{N}} = \sqrt{\frac{\sum_{i=1}^{N}x_i^2}{N}-\mu^2} = \sqrt{\frac{\sum_{i=1}^{N}x_i^2}{N}-\left(\frac{\sum_{i=1}^{N}x_i}{N}\right)^2}$$

$$S = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}} = \sqrt{\frac{\sum_{i=1}^{n}x_i^2-n\bar{x}^2}{n-1}} = \sqrt{\frac{n\sum_{i=1}^{n}x_i^2-\left(\sum_{i=1}^{n}x_i\right)^2}{n(n-1)}}$$

การวัดการกระจาย

เมื่อ	σ	แทน	ส่วนเบี่ยงเบนมาตรฐานของประชากร
	S	แทน	ส่วนเบี่ยงเบนมาตรฐานของกลุ่มตัวอย่าง
	μ	แทน	ค่าเฉลี่ยเลขคณิตของประชากร
	$\bar{X}$	แทน	ค่าเฉลี่ยเลขคณิตของกลุ่มตัวอย่าง
	X_i	แทน	ค่าหรือคะแนนต่างๆ ของข้อมูลที่ i
	N	แทน	จำนวนข้อมูลของประชากร
	n	แทน	จำนวนข้อมูลของกลุ่มตัวอย่าง

การวัดการกระจาย

1.4.2 ส่วนเบี่ยงเบนมาตรฐานกรณีข้อมูลแจกแจงความถี่ สามารถหาคำนวณได้จากสูตรดังนี้

$$\sigma = \sqrt{\frac{\sum_{i=1}^{k} f_i\left(X_i - \mu\right)^2}{N}} = \sqrt{\frac{\sum_{i=1}^{k} f_i X_i^2}{N} - \left(\frac{\sum_{i=1}^{k} f_i X_i}{N}\right)^2}$$

$$S = \sqrt{\frac{\sum f_i(X_i - \bar{X})^2}{n-1}} = \sqrt{\frac{n\sum f_i X_i^2 - \left(\sum f_i X_i\right)^2}{n(n-1)}}$$

การวัดการกระจาย

1.5 ความแปรปรวน (variance) เป็นค่ากำลังสองของส่วนเบี่ยงเบนมาตรฐาน ดังนั้นการคำนวณค่าความแปรปรวนสามารถคำนวณเช่นเดียวกับค่าส่วนเบี่ยงเบนมาตรฐาน แต่ไม่ต้องถอดรากที่สอง นั่นคือเท่ากับค่าส่วนเบี่ยงเบนมาตรฐานยกกำลังสอง ในกรณีประชากรจะแทนด้วยสัญลักษณ์ และความแปรปรวนของกลุ่มตัวอย่างแทนด้วยสัญลักษณ์ คำนวณได้ดังนี้

การวัดการกระจาย

1.5.1 ความแปรปรวนของประชากรและกลุ่มตัวอย่างกรณีข้อมูลไม่แจกแจงความถี่ สามารถหาได้ดังนี้

$$\sigma^2 = \frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N} = \frac{\sum_{i=1}^{N}x_i^2}{N} - \left(\frac{\sum_{i=1}^{N}x_i}{N}\right)^2$$

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

การวัดการกระจาย

1.5.2 ความแปรปรวนของประชากรและกลุ่มตัวอย่างกรณีข้อมูลแจกแจงความถี่ สามารถหาได้ดังนี้

$$\sigma^2 = \frac{\sum_{i=1}^{k} f_i (X_i - \mu)^2}{N} = \frac{\sum_{i=1}^{k} f_i X_i^2}{N} - \left(\frac{\sum_{i=1}^{k} f_i X_i}{N} \right)^2$$

$$S^2 = \frac{\sum f_i (X_i - \bar{X})^2}{n - 1} = \frac{n \sum f_i X_i^2 - \left(\sum f_i X_i \right)^2}{n(n - 1)}$$

การวัดการกระจาย

2. การวัดการกระจายสัมพัทธ์ (relative variation) เป็นการหาค่าเพื่อเปรียบเทียบ การกระจายระหว่างข้อมูลมากกว่าหนึ่งชุด ซึ่งเหมาะสำหรับข้อมูลที่แตกต่างกันมากจึงหาอัตราส่วนของค่าที่ได้จากการวัดการกระจายสัมบูรณ์กับค่ากลางของข้อมูลชุดนั้นๆ ซึ่งมีอยู่ 4 ชนิด ได้แก่

สัมประสิทธิ์ของพิสัย (C.R.)

สัมประสิทธิ์ของส่วนเบี่ยงเบนควอร์ไทล์ (C.Q.D.)

สัมประสิทธิ์ของส่วนเบี่ยงเบนเฉลี่ย (C.M.D.)

สัมประสิทธิ์ของการแปรผัน (C.V.)

การวัดการกระจาย

สัมประสิทธิ์ของพิสัย (C.R.) $= \dfrac{X_{max} - X_{min}}{X_{max} + X_{min}}$

สัมประสิทธิ์ของส่วนเบี่ยงเบนควอร์ไทล์ (C.Q.D.) $= \dfrac{Q_3 - Q_1}{Q_3 + Q_1}$

สัมประสิทธิ์ของส่วนเบี่ยงเบนเฉลี่ย (C.M.D.) $= \dfrac{M.D.}{\bar{x}}$

สัมประสิทธิ์ของการแปรผัน (C.V.) $= \dfrac{S}{\bar{X}} \times 100\%$

ตาราง การเลือกใช้การวัดการกระจาย

การวัดการกระจาย	ระดับการวัด			
	นามบัญญัติ	เรียงอันดับ	อันตรภาค	อัตราส่วน
พิสัย	-	-	✓	✓
ส่วนเบี่ยงเบนควอร์ไทล์	-	✓	✓	✓
ส่วนเบี่ยงเบนมาตรฐาน	-	-	✓	✓
ความแปรปรวน	-	-	✓	✓

การวิเคราะห์ข้อมูลเบื้องต้น โดยใช้โปรแกรมคอมพิวเตอร์

โดยใช้โปรแกรมคอมพิวเตอร์ SPSS มีขั้นตอนดังนี้

ขั้นที่ 1 กรอกข้อมูลในโปรแกรม

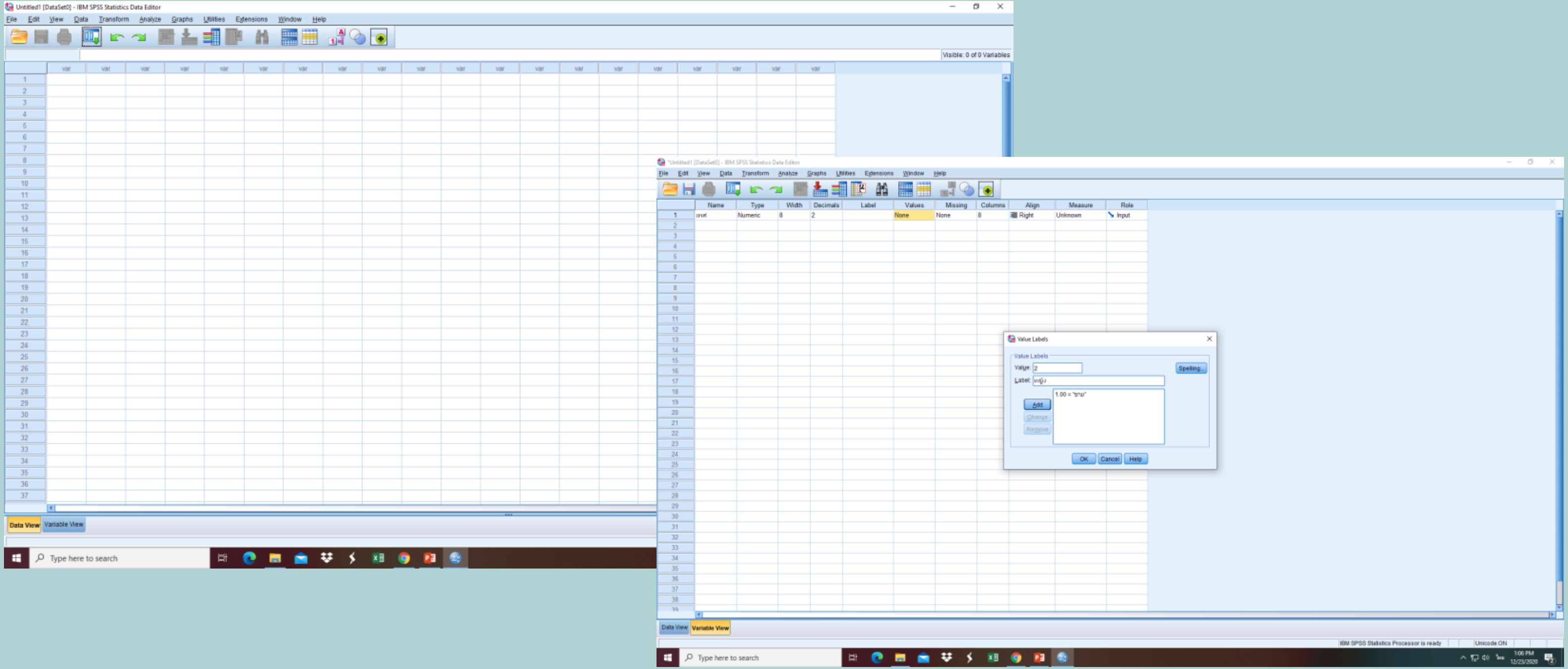
ขั้นที่ 2 คลิกเลือกคำสั่ง Analyze → Descriptive Statistics → Frequencies

ขั้นที่ 3 เลือกตัวแปรที่ต้องการวิเคราะห์ในช่อง Variable(s)

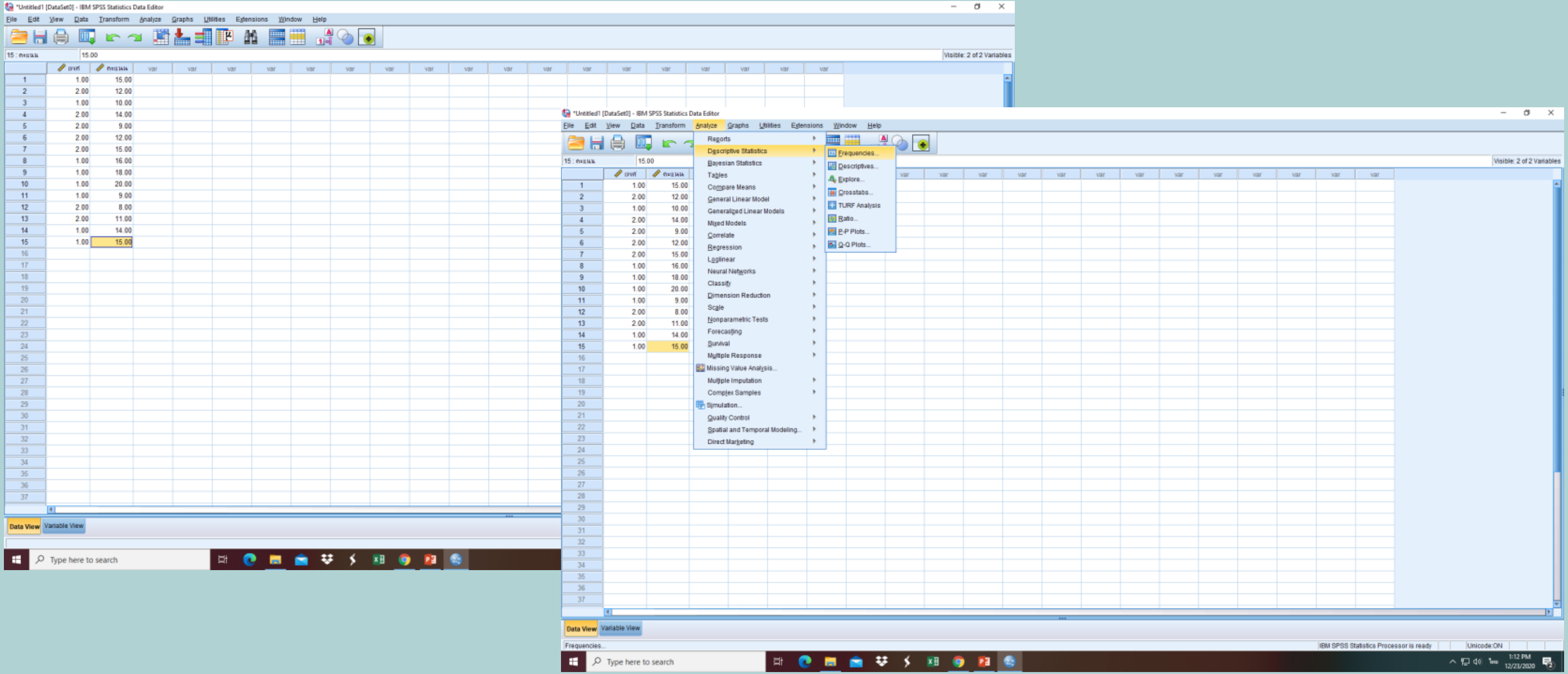
ขั้นที่ 4 คลิก Statistics... เพื่อวิเคราะห์คำนวณค่า ได้แก่ ค่าเฉลี่ย ค่ามัธยฐาน ค่าฐานนิยม และการวัดการกระจาย แล้วคลิก Continue

ขั้นตอนที่ 5 คลิก OK

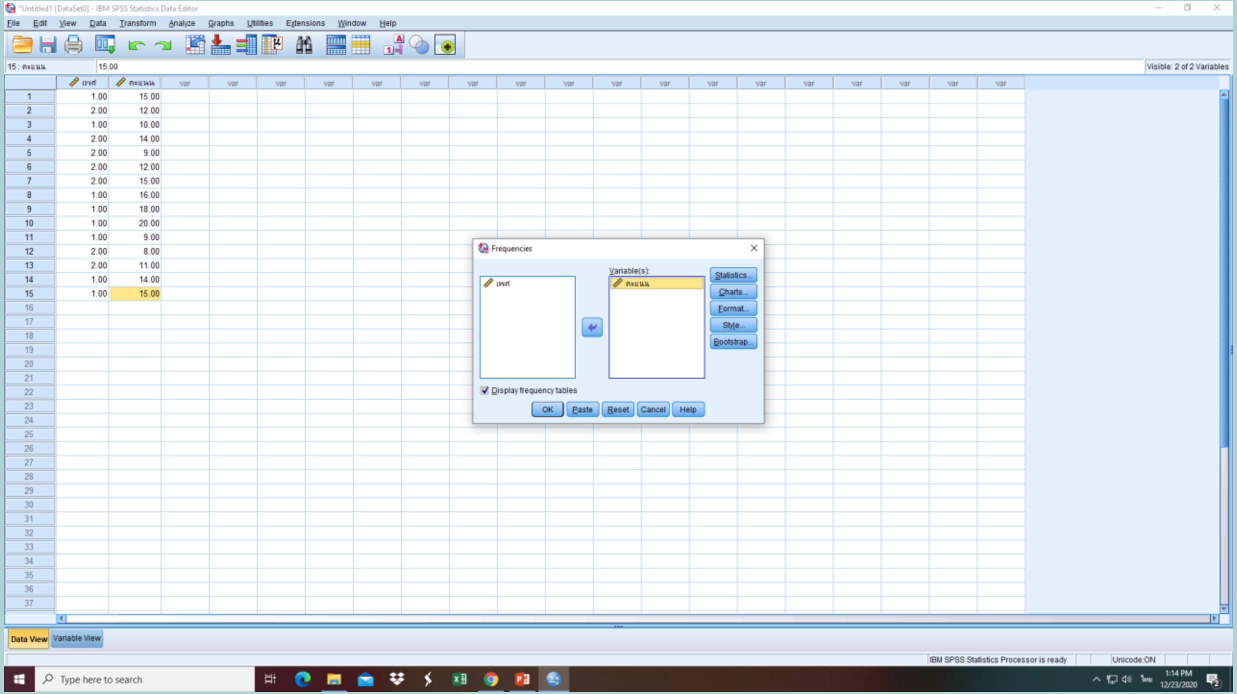
ขั้นที่ 1 กรอกข้อมูลในโปรแกรม



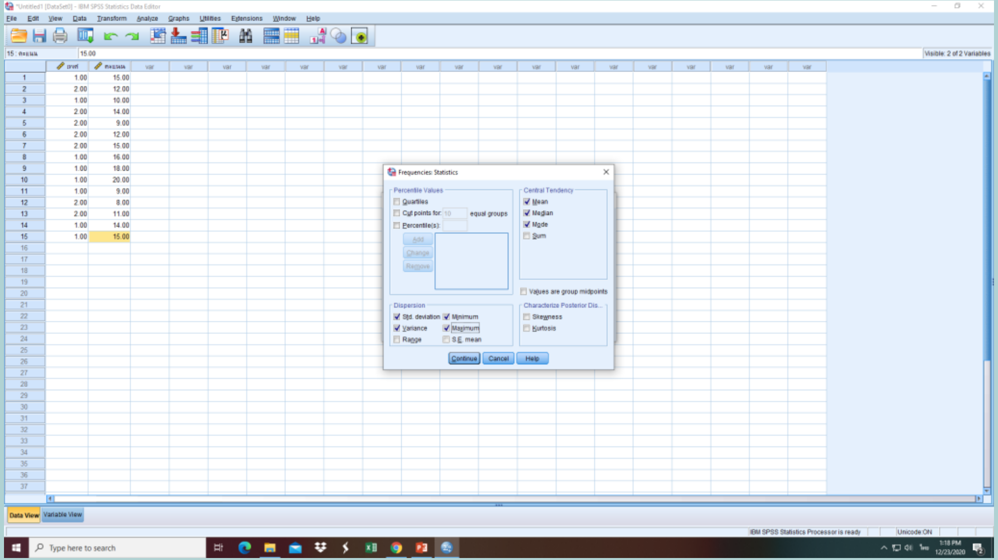
ขั้นที่ 2 คลิกเลือกคำสั่ง Analyze ➡ Descriptive Statistics ➡ Frequencies



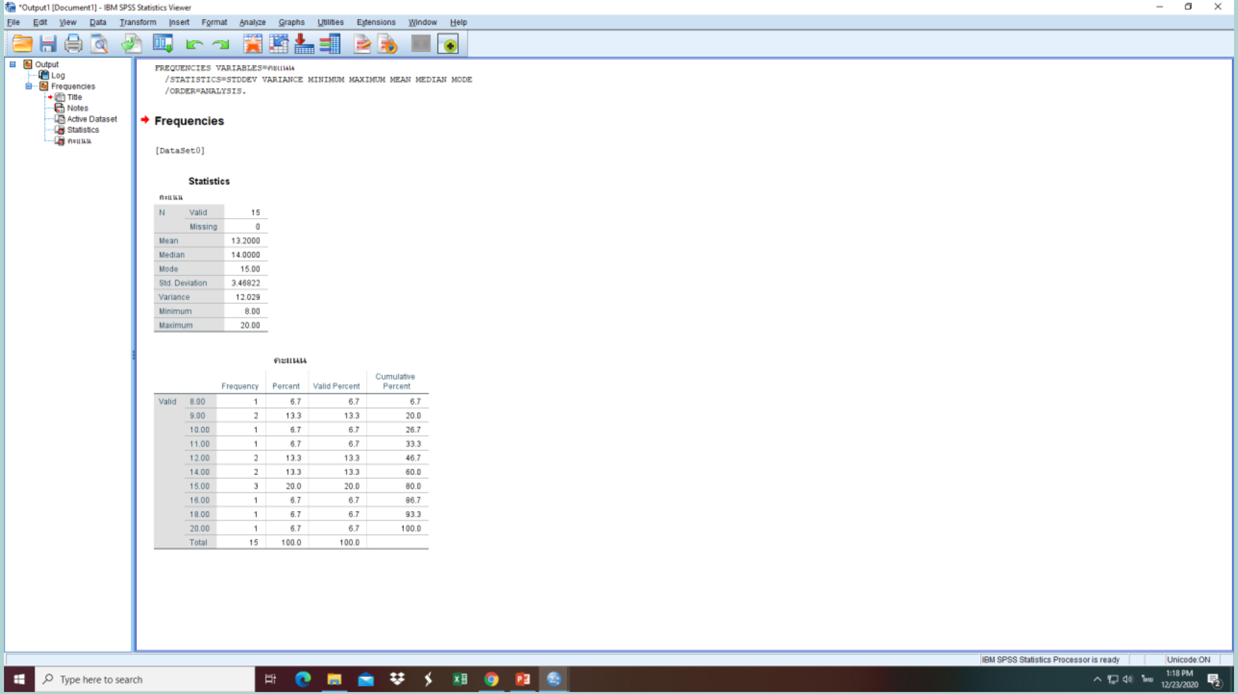
ขั้นที่ 3 เลือกตัวแปรที่ต้องการวิเคราะห์ในช่อง Variable(s)



ขั้นที่ 4 คลิก Statistics... เพื่อวิเคราะห์คำนวณค่า ได้แก่ ค่าเฉลี่ย ค่ามัธยฐาน ค่าฐานนิยม และการวัดการกระจาย แล้วคลิก Continue



ขั้นตอนที่ 5 คลิก OK




```
FREQUENCIES VARIABLES=คะแนน
  /STATISTICS=STDDEV VARIANCE MINIMUM MAXIMUM MEAN MEDIAN MODE
  /ORDER=ANALYSIS.
```

Frequencies

[DataSet0]

Statistics

คะแนน

N	Valid	15
	Missing	0
Mean		13.2000
Median		14.0000
Mode		15.00
Std. Deviation		3.46822
Variance		12.029
Minimum		8.00
Maximum		20.00

คะแนน

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	8.00	1	6.7	6.7	6.7
	9.00	2	13.3	13.3	20.0
	10.00	1	6.7	6.7	26.7
	11.00	1	6.7	6.7	33.3
	12.00	2	13.3	13.3	46.7
	14.00	2	13.3	13.3	60.0
	15.00	3	20.0	20.0	80.0
	16.00	1	6.7	6.7	86.7
	18.00	1	6.7	6.7	93.3
	20.00	1	6.7	6.7	100.0
	Total	15	100.0	100.0	

ตัวอย่างงานวิจัยทางการศึกษาที่ใช้การวิเคราะห์ข้อมูลเบื้องต้น

อุษา จันทร (2552) ทำวิจัยเรื่อง “การพัฒนากิจกรรมการเรียนรู้คณิตศาสตร์ เรื่อง การหาร สำหรับนักเรียนชั้นประถมศึกษาปีที่ 3 โดยใช้รูปแบบแนวคิดของทฤษฎีคอนสตรัคติวิสต์ ที่เน้นทักษะ/กระบวนการทางคณิตศาสตร์”

วัตถุประสงค์การวิจัย: เพื่อศึกษาผลการพัฒนาทักษะ/กระบวนการทางคณิตศาสตร์ ของนักเรียนชั้นประถมศึกษาปีที่ 3 โรงเรียนบ้านหัวบึง อำเภอน้ำพอง จังหวัดขอนแก่น ที่เกิดจากผลการพัฒนากิจกรรมการเรียนรู้คณิตศาสตร์ เรื่อง การหาร โดยใช้รูปแบบแนวคิดของทฤษฎีคอนสตรัคติวิสต์ที่เน้นทักษะ/กระบวนการทางคณิตศาสตร์

ตัวอย่างงานวิจัยทางการศึกษาที่ใช้การวิเคราะห์ข้อมูลเบื้องต้น

สถิติที่ใช้ในการวิเคราะห์ข้อมูล : ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน และการหาร้อยละ

ผลการวิเคราะห์ข้อมูล :

ตารางผลการทดสอบท้ายวงจร

วงจรที่	จำนวนนักเรียนทั้งหมด	คะแนน							จำนวนนักเรียนที่ผ่านเกณฑ์	
		เต็ม	ผ่านเกณฑ์	สูงสุด	ต่ำสุด	ค่าเฉลี่ย	ร้อยละ	ส่วนเบี่ยงเบนมาตรฐาน	จำนวน (คน)	ร้อยละ
1	13	20	14	20	9	16.69	83.46	3.33	11	84.62

จากตารางพบว่า ผลการทดสอบท้ายวงจรที่ 1 นักเรียนได้คะแนนเฉลี่ยร้อยละ 83.46 และมีนักเรียนได้คะแนนผ่านเกณฑ์ร้อยละ 70 ของคะแนนเต็ม จำนวน 11 คน คิดเป็นร้อยละ 84.62

ตัวอย่างงานวิจัยทางการศึกษาที่ใช้การวิเคราะห์ข้อมูลเบื้องต้น

ธนวัฒน์ ศรีศิริวัฒน์ (2562) ทำวิจัยเรื่อง “การศึกษาผลสัมฤทธิ์ทางการเรียนเรื่อง ความน่าจะเป็น ของนักศึกษาชั้นปีที่ 1 สาขาวิชาคณิตศาสตร์ โดยการจัดกิจกรรมการเรียนรู้แบบ 4MAT ”

วัตถุประสงค์การวิจัย: เพื่อเปรียบเทียบผลสัมฤทธิ์ทางการเรียนเรื่อง ความน่าจะเป็น ของนักศึกษาชั้นปีที่ 1 สาขาวิชาคณิตศาสตร์ คณะครุศาสตร์ มหาวิทยาลัยราชภัฏสวนสุนันทา หลังได้รับการจัดกิจกรรมการเรียนรู้แบบ 4 MAT กับเกณฑ์ร้อยละ 60

ตัวอย่างงานวิจัยทางการศึกษาที่ใช้การวิเคราะห์ข้อมูลเบื้องต้น

สถิติที่ใช้ในการวิเคราะห์ข้อมูล : ร้อยละ ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน และการทดสอบค่าที

ผลการวิเคราะห์ข้อมูล

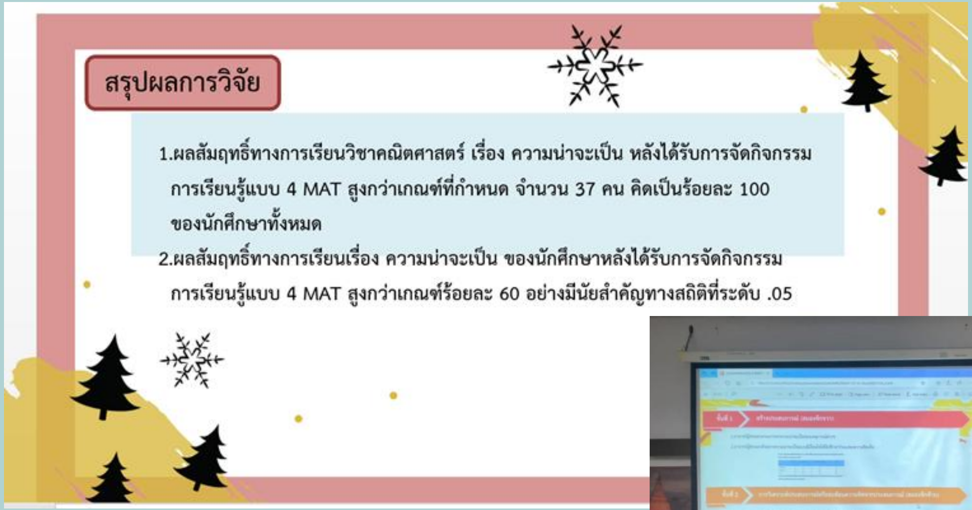
ตารางที่ 1 แสดงจำนวนและร้อยละของนักเรียนที่มีผลสัมฤทธิ์ทางการเรียน เรื่อง ความน่าจะเป็น สูงกว่าเกณฑ์ (ร้อยละ 60)

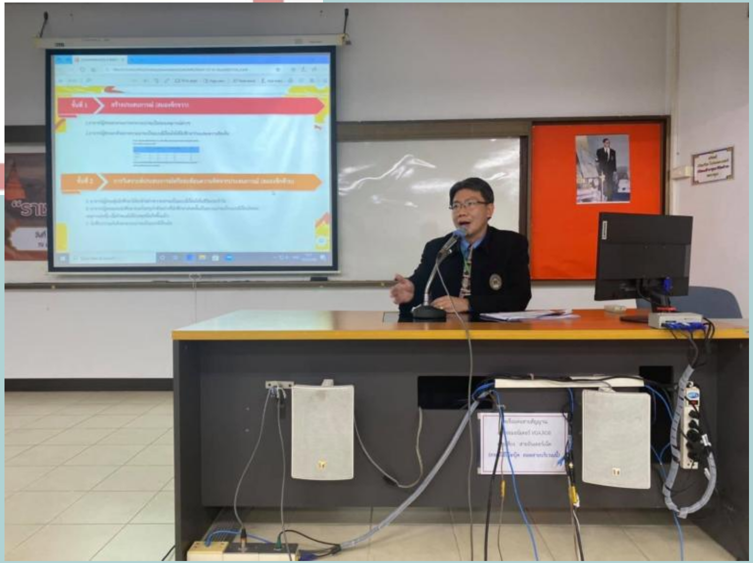
สภาพการณ์	จำนวนนักศึกษา	จำนวนนักศึกษาที่ได้คะแนนสูงกว่าเกณฑ์	ร้อยละของจำนวนนักศึกษาที่ได้คะแนนสูงกว่าเกณฑ์
หลังเรียน	37	37	100

ตารางที่ 2 แสดงผลการวิเคราะห์ผลสัมฤทธิ์ทางการเรียน เรื่อง ความน่าจะเป็น หลังได้รับการจัดกิจกรรมการเรียนรู้แบบ 4 MAT

สภาพการณ์	จำนวนนักศึกษา	คะแนนเต็ม	คะแนนร้อยละ60	$\bar{x}$	S.D.	t	Sig
หลังเรียน	37	20	12	13.89	.70	16.47*	.00

**มีนัยสำคัญทางสถิติที่ระดับ .05


สรุปผลการวิจัย
1.ผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ เรื่อง ความน่าจะเป็น หลังได้รับการจัดกิจกรรม การเรียนรู้แบบ 4 MAT สูงกว่าเกณฑ์ที่กำหนด จำนวน 37 คน คิดเป็นร้อยละ 100 ของนักศึกษาทั้งหมด
2.ผลสัมฤทธิ์ทางการเรียนเรื่อง ความน่าจะเป็น ของนักศึกษาหลังได้รับการจัดกิจกรรม การเรียนรู้แบบ 4 MAT สูงกว่าเกณฑ์ร้อยละ 60 อย่างมีนัยสำคัญทางสถิติที่ระดับ .05



อ้างอิง

ธนวัฒน์ ศรีศิริวัฒน์. (2560). สถิติเพื่อการวิจัย. กรุงเทพฯ : มหาวิทยาลัยราชภัฏสวนสุนันทา.