บทที่ 3 สถิติประยุกต์สำหรับการจัดการคุณภาพ

<u>**ตัวอย่างที่2**</u> จากข้อมูลผลการตรวจสอบอายุการใช้งานของเครื่องมือสำหรับงานช่างชนิดหนึ่ง จำนวน 50 อัน ได้ข้อมูลดังนี้ (หน่วยเป็นสัปดาห์และเศษของสัปดาห์ปัดขึ้น)

33	47	72	63	52	52	35	50	37	43	53	43	52	44
42	31	36	48	43	27	58	62	49	34	48	53	39	45
34	59	34	66	40	59	36	41	35	36	62	34	38	28
43	50	30	43	32	45	59	55						

ให้นักศึกษาสร้างฮีสโทแกรมความถี่สัมพัทธ์เพื่ออธิบายชุดข้อมูลดังกล่าว

วิธีทำ จากข้อมูลที่ได้มาสามารถสร้างฮีสโทแกรมความถี่สัมพัทธ์ด้วยขั้นตอนและวิธีการดังนี้

ขั้นตอนที่ 1 เลือกกำหนดจำนวนช่วงของข้อมูลหรือจำนวนอันตรภาคชั้น **= 6 ช่วง**

ขั้นตอนที่ 2 คำนวณหาความกว้างของแต่ละอันตรภาคชั้นโดยประมาณ ดังนี้

= (72 – 27) / 6 = 7.50

จะสังเกตได้ว่าค่าที่คำนวณได้นี้เป็นตัวเลขที่เป็นจุดทศนิยมให้ทำหารปัดเศษขึ้นเพื่อให้มีค่าเป็นตัวเลขจำนวนเต็ม ดังนั้นความกว้างของแต่ละอันตรภาคชั้นจะมีค่า = 8 นั่นเอง

ขั้นตอนที่ 3 กำหนดช่วงของข้อมูลในแต่ละอันตรภาคชั้น โดยเริ่มจากข้อมูลค่าที่น้อยที่สุดในชุดข้อมูล

คือ 27 อย่างไรก็ตาม เราสามารถเลือกค่าเริ่มต้นให้เป็นค่าอื่นที่ใกล้เคียงได้ โดยต้องมีค่าน้อยกว่า 27 ซึ่งในที่นี้ได้เลือกที่จะกำหนดให้ค่าต่ำสุดของการสร้างฮีสโทแกรม **คือ 25**

ดังนั้น สามารถกำหนดช่วงข้อมูลในแต่ละอันตรภาคชั้นได้ ดังนี้

ตารางที่ 3.2 การแบ่งช่วงของข้อมูลออกเป็น 6 อันตรภาคชั้น

อันตรภาคชั้นที่	ช่วงของข้อมูล
1	25-32
2	33-40
3	41-48
4	49-56
5	57-64
6	65-73

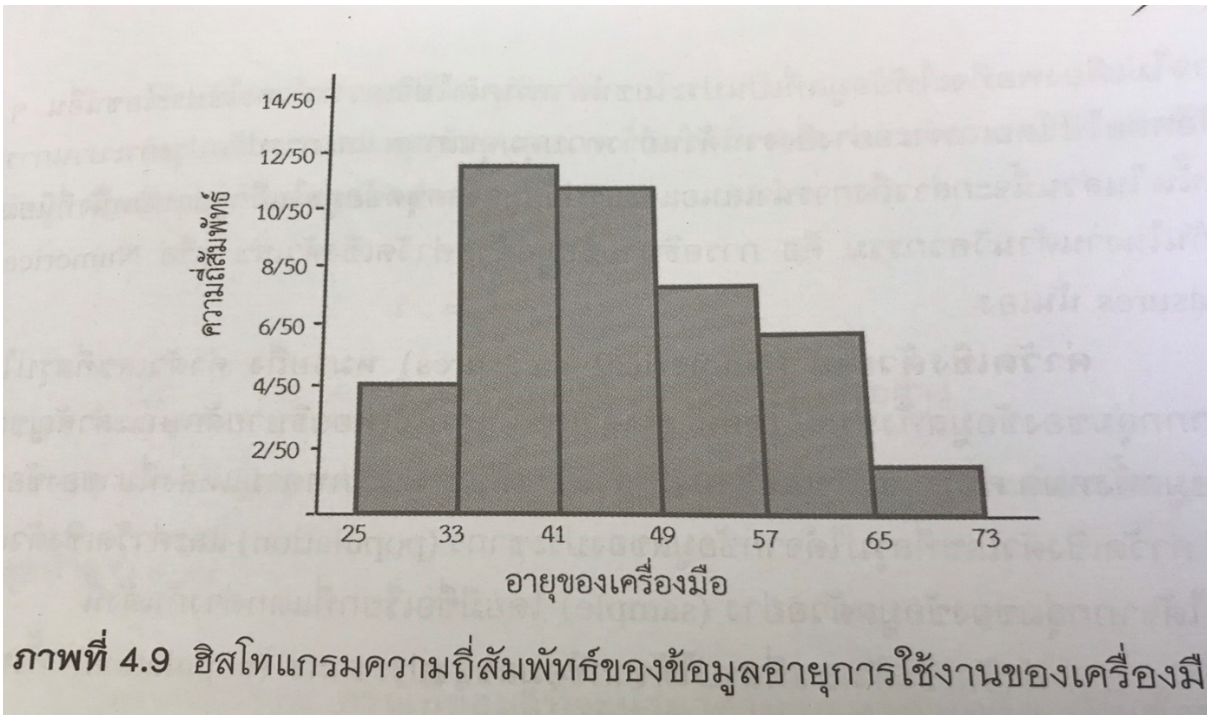
จะเห็นได้ว่าเมื่อดำเนินการตามขั้นตอนที่ 1จนถึงขั้นที่ 3 นี้แล้วอันตรภาคชั้นทั้งหมดที่กำหนดขึ้นจะครอบคลุมค่าที่น้อยที่สุดถึงมากที่สุดของชุดข้อมูล โดยมีจำนวนอันตรภาคชั้นเท่าที่กำหนดไว้ในขั้นตอนที่ 1

ขั้นตอนที่ 4 สร้างตารางสถิติ ด้วยจำนวนช่วงข้อมูลทั้งสิ้น 6 ช่วง แต่ละช่วงครอบคลุมค่าของข้อมูล 8 ค่าและค่าต่ำสุดของอันตรภาคชั้นแรกคือ 25 โดยใช้ข้อมูลทั้งหมดที่ให้มาจากนั้นทำการตรวจสอบ นับความถี่ของข้อมูลในแต่ละช่วง และคำนวณหาค่าความถี่สัมพัทธ์ ซึ่งได้ผลดังนี้

ตารางที่ 4.3 การนับความถี่ และคำนวณหาความถี่สัมพัทธ์ของแต่ละอันตรภาคชั้น

อายุของเครื่องมือ (สัปดาห์)	**การนับ**	**ความถี่**	**ความถี่สัมพัทธ์**
25-32	IIIII	5	5/50 = 0.10
33-40	IIIII IIIII IIII	14	14/50 = 0.28
41-48	IIIII IIiII III	13	13/50 = 0.26
49-56	IIIII IIII	9	9/50 = 0.18
57-64	IIIII II	7	7/50 = 0.14
65-73	II	2	2/50 = 0.04

ขั้นตอนที่ 5 สร้างกราฟแท่งเพื่อแจกแจงค่าความถี่สัมพัทธ์ของข้อมูลในแต่ละช่วง โดยให้แกน x คือช่วงอายุของอุปกรณ์นี้ และแกน y คือความถี่สัมพัทธ์ที่ได้จากขั้นตอนที่ 4 ซึ่งจะทำให้ได้ฮีสโทแกรมความถี่สัมพัทธ์ที่สมบรูณ์



ภาพที่ 3.9 **ฮี**สโทแกรมความถี่สัมพัทธ์ของข้อมูลอายุการใช้งานของเครื่องมือ

จากฮีสโทแกรมความถี่สัมพัทธ์ที่สร้างขึ้น สามารถนำมาใช้วิเคราะห์ข้อมูลและตอบคำถามที่สำคัญต่าง ๆ ได้ เช่น

1) สัดส่วนส่วนของเครื่องมือนี้ที่มีอายุใช้งานน้อยกว่า 41 สัปดาห์ มีค่าเท่าใด

$$= (14+5)/50 = 19/50 = 0.38$$

2) ค่าความน่าจะเป็นที่สุ่มเลือกเครื่องมือนี้ขึ้นมา 1 อัน แล้วพบว่าเครื่องมืออันที่เลือกมีอายุการใช้งานตั้งแต่ 49 สัปดาห์ขึ้นไป มีค่าเป็นเท่าใด

$$= (9+7+2)/50 = 17/50 = 0.34$$

จากตัวอย่างจะเห็นได้ว่าเราสามารถนำฮีสโทแกรมความถี่สัมพทธ์ไปใช้ประโยชน์ในการวิเคราะห์หาความน่าจะเป็นของการเกิดเหตุการณ์ใด ๆ ได้ โดยเรียกว่ากราฟแสดงการแจกแจงความน่าจะเป็น ซึ่งจะได้กล่าวถึงในบทที่ 4 ต่อไป

3.1.3 การนำเสนอข้อมูลด้านคุณภาพด้วยค่าเชิงตัวเลข

ถึงแม้ว่าการอธิบายข้อมูลโดยใช้กราฟเป็นวิธีการที่ช่วยให้ผู้อ่านสามารถเข้าถึงและเข้าใจข้อมูลโดยภาพรวมได้อย่างชัดเจนและรวดเร็วก็ตาม แต่สำหรับงานด้านวิทยาศาสตร์และวิศวกรรมศาสตร์นั้น บางครั้งการอธิบายข้อมูลด้วยกราฟหรือแผนภาพอย่างเดียวอาจจะไม่เพียงพอที่จะให้ข้อมูลที่เป็นประโยชน์สำหรับนำไปวิเคราะห์และใช้ประโยชน์อื่น ๆได้อีกต่อไป โดยเฉพาะอย่างยิ่งงานด้านการควบคุมคุณภาพ และการปรับปรุงกระบวนการ ดังนั้นในส่วนนี้จะกล่าวถึงการนำเสนอและสรุปข้อมูลจากชุดข้อมูลในอีกรูปแบบหนึ่งที่นิยมใช้กันในงานวิศวกรรม คือ การอธิบายข้อมูลด้วยค่าวัดเชิงตัวเลข หรือ Numerical measures นั่นเอง

ค่าวัดเชิงตัวเลข (Numerical measures) หมายถึงค่าตัวเลขที่สรุปได้มาจากกลุ่มของข้อมูลทั้งหมด โดยค่าเชิงตัวเลขนั้นถูกใช้เพื่ออธิบายลักษณะสำคัญของชุด

ข้อมูลทั้งหมด ค่าวัดเชิงตัวเลขสามารถจำแนกเป็น 2 ประเภทตามแหล่งที่มาของข้อมูล ได้แก่ ค่าวัดเชิงตัวเลขที่สรุปได้จากข้อมูลของประชากร (population) และค่าวัดเชิงตัวเลขที่สรุปได้จากกลุ่มของข้อมูลตัวอย่าง (sample) โดยมีชื่อเรียกที่แตกต่างกันดังนี้

1) ค่าวัดเชิงตัวเลขที่ที่สรุปได้จากข้อมูลของประชากร (population) ทั้งหมดที่ทำการศึกษา จะเรียกค่านั้นว่าค่า “**ค่าพารามิเตอร์** (parameter)”
2) ค่าวัดเชิงตัวเลขที่สรุปได้จากข้อมูลของกลุ่มตัวอย่าง(sample) ซึ่งคัดเลือกมาเพียงบางส่วน จากกลุ่มประชากรทั้งหมดที่ทำการศึกษา เรียกค่านั้นว่า “**ค่าสถิติ** (statisitics)

สำหรับการอธิบายข้อมูลด้านคุณภาพด้วยค่าวัดเชิงตัวเลขนั้น ค่าวัดที่สำคัญและจำเป็นต้องใช้ในการอธิบายลักษณธข้อมูล ประกอบด้วยค่าวัด 2 ประเภท ได้แก่

1. ค่าแสดงแนวโน้มศูนย์กลางหรือเป็นตัวแทนของชุดข้อมูล
2. ค่าวัดแสดงการกระจายของข้อมูล ซึ่งอธิบายรายละเอียดได้ดังนี้

❖ 1. ค่าแนวโน้มศูนย์กลางของข้อมูลด้านคุณภาพ

เป็นค่าวัดที่ใช้เป็นตัวแทนขอชุดข้อมูล ซึ่งมีลักษณะกระจายไปตามแกนนอน หรือแกน X หรือระบุกำหนดตำแหน่งที่มีแนวโน้มและมีจุดศูนย์กลางของชุดข้อมูลนั้น

สำหรับค่าวัดที่ใช้แสดงถึงแนวโน้มศูนย์กลางของข้อมูลที่นิยมใช้กันทั่วไปมี 3 แบบ ได้แก่ ค่าเฉลี่ย (mean) ค่ามัธยฐาน (median) และค่าฐานนิยม(mode)อย่างไรก็ตาม สำหรับงานด้านควบคุมคุณภาพ ค่าตัวแทนของชุดข้อมูลที่นิยมใช้กันคือ ค่าเฉลี่ยของข้อมูล

ค่าเฉลี่ยของข้อมูลด้านคุณภาพ

คือ ผลรวมของข้อมูลด้านคุณภาพที่วัดได้ทั้งหมด แล้วนำมาหารด้วยจำนวนข้อมูล ซึ่งอธิบายได้ดังสูตร ต่อไปนี้

$$\bar{X} = \frac{\sum_{i=1}^{n} Xi}{n}$$; เมื่อ n = จำนวนข้อมูลทั้งหมดของกลุ่มตัวอย่าง

อย่างไรก็ตาม กรณีที่ข้อมูลที่จะนำมาคำนวณค่าเฉลี่ย เป็นข้อมูลของประชากรทั้งหมด จะใช้ชื่อเรียกแตกต่างกันดังนี้

$$\boldsymbol{\mu} = \frac{\sum_{i=1}^{n} \boldsymbol{Xi}}{\boldsymbol{N}}$$; เมื่อ N = จำนวนข้อมูลทั้งหมดขงประชากร

ตัวอย่าง 3.7 จากผลการสุ่มตรวจสอบน้ำหนักของขนมชนิดหนึ่งที่มีการวางขายในตลาด จำนวน 30 ตัวอย่าง ได้ผลดังนี้

29.88	*30.04*	*29.71*	29.87	30.05	30.03	30.02	30.01	29.70	30.05	30.01	29.50	29.95
29.36	30.34	30.02	30.00	29.41	29.80	30.07	29.83	29.76	29.90	30.04	29.78	30.03
29.60	30.02	30.00	29.71									

จากข้อมูลที่ให้มา ให้คำนวณหาค่าเฉลี่ยของน้ำหนักขนม จำนวน 30 ชิ้นนี้

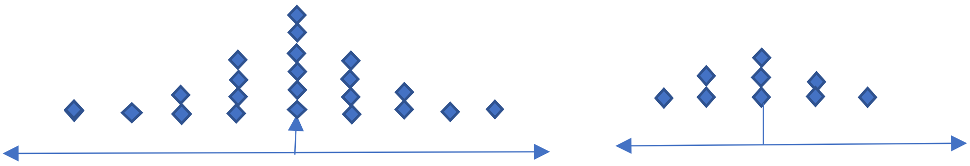
วิธีทำ ค่าเฉลี่ยของข้อมูลในโจทย์ สามารถคำนวณ ได้ดังนี้

$$\bar{X} = \frac{\sum_{i=1}^{n} Xi}{30} = \frac{29.88+29.36+30.04+\cdots.29.71}{30}$$

หมายเหตุ; ในกรณีที่ค่าเฉลี่ยนี้ หามาจากข้อมูลที่มาจากประชากรทั้งหมด เราจะเรียกค่าเฉลี่ยนี้ว่า μ

2. ค่าแสดงการกระจายของข้อมูลด้านคุณภาพ

เป็นค่าวัดที่ใช้แสดงระดับการกระจายตัว *(spread)* **ของข้อมูลด้านคุณภาพทั้งหมดในชุดข้อมูลซึ่งพิจารณาได้จาก** *dot plot***ด้านล่างนี้**



ภาพที่ 3.10 การกระจายตัวของข้อมูล 2 ชุดที่แตกต่างกัน

จากภาพจะเห็นได้ว่าข้อมูล 2 ชุดนี้มีระดับการกระจายตัวที่แตกต่างกัน ซึ่งการอธิบายระดับการกระจายข้อมูลออกมาเป็นค่าเชิงตัวเลขเพื่อเปรียบเทียบกันนี้ ค่าวัดที่นิยมใช้ในปัจจุบันมี 2 ค่า คือ ค่าพิสัย และความแปรปรวนของข้อมูล สำหรับความหมาย วิธีการคำนวนและตัวอย่างการคำนวนค่าวัดการกระจายทั้ง 2แบบนี้

1) **พิสัย** *(Range)*

พิสัยของชุดข้อมูล คือ ผลต่างระหว่างข้อมูลที่มีค่ามากที่สุด (Xmax) ***กับข้อมูลที่มีค่าน้อยที่สุด*** (Xmin)

R = Xmax – Xmin

ห้อง 2 ค่าการกระจายคือ 26-20 = 6 ***ค่าเฉลี่ย คือ 26+20/2*** = 23

ห้อง 3 ค่าการกระจายคือ 28 -26 = 2 ***ค่าเฉลี่ย คือ 28+26/2*** = 27

*ค่าพิสัย*จัดเป็นค่าแสดงการกระจายของข้อมูลที่สามารถคำนวนหาได้ง่ายและรวดเร็ว โดยใช้ค่าเพียง 2 ค่า ในชุดข้อมูลเท่านั้น ถ้าค่าพิสัยมาก แสดงว่าความแตกต่างระหว่างค่าที่มากที่สุด และค่าที่น้อยที่สุดในชุดข้อมูลมาก ส่วนค่าพิสัยที่น้อยกว่า แสดงว่าผลต่างนี้มีค่าน้อย

กว่า ซึ่งบ่งบอกว่าการกระจายของข้อมูลชุดนั้นน้อยกว่าชุดข้อมูลที่มรค่าพิสัยมากนั่นเอง

ตัวอย่างที่ **3.8** ฝ่ายตรวจสอบคุณภาพ ทำการบันทึกจำนวนรอยตำหนิที่พบบนม้วนผ้า จำนวน 8 ม้วนได้ข้อมูลดังนี้

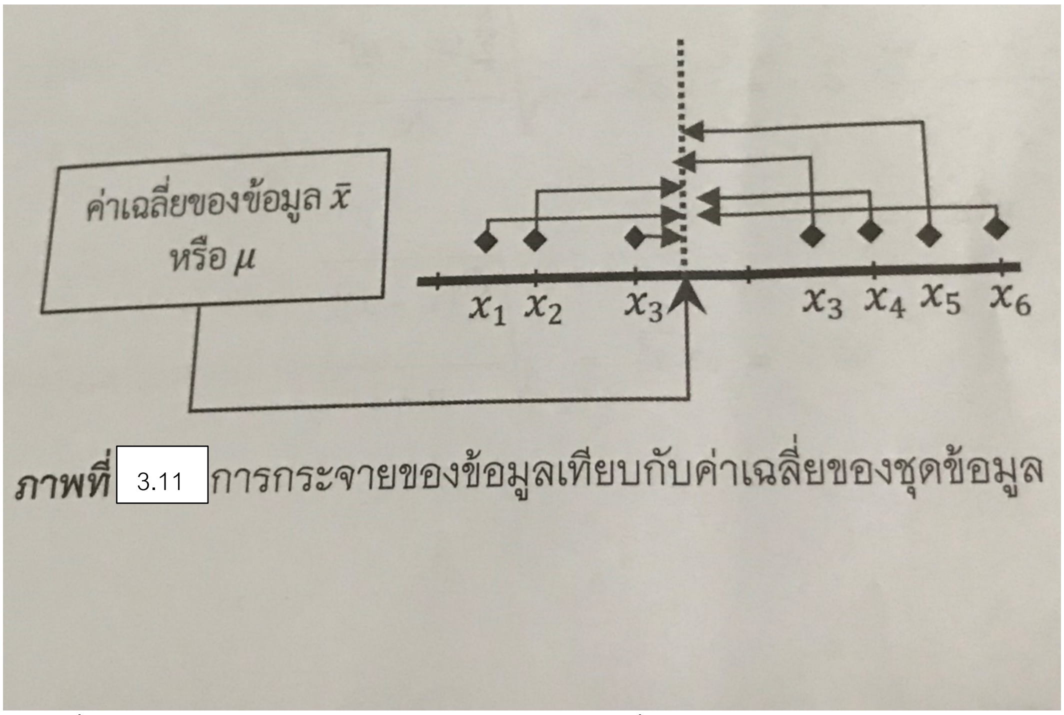
6 , 7, 11 , 7 , 8 , 4 , 9 , 12

วิธีทำ **จากข้อมูลที่ได้มา จะได้** *Xmax = 12 , Xmin=4*

เพราะฉะนั้นค่าพิสัย (R) = 12- 4 = 8

2) ***ความแปรปรวน*** (Variance) ***ของข้อมูล***

ความแปรปรวน คือ *ค่าแสดงระดับการกระจายหรือความผันแปรของข้อมูลในอีกรูปแบบหนึ่ง ซึ่งหมายถึงค่าเฉลี่ยของความเบี่ยงเบนยกกำลังสองของข้อมูลแต่ละค่าเทียบกับค่าเฉลี่ยของชุดข้อมูลนั้น โดยที่มาของสูตรการคำนวณสามารถอธิบายภาพ ดังนี้*



ภาพที่ 3.11 การกระจายของข้อมูลเทียบกับค่าเฉลี่ยของข้อมูล

จากภาพเมื่อนำข้อมูลแต่ละค่ามาพล็อตในลักษณะของ *dot plot* โดยมีเส้นประแทนค่าเฉลี่ย ซึ่งมีค่าเท่ากับ *X* (กรณีที่ข้อมูลในชุดนั้นเป็นข้อมูลจากกลุ่มตัวอย่าง) หรือมีค่าเท่ากับ μ (กรณีของชุดข้อมูลเป็นกลุ่มประชากร) จากความหมายของความแปรปรวน จะได้สูตรคำนวณค่าความแปรปรวน ของข้อมูลดังนี้

ค่าความแปรปรวนของประชากร *N*ซึ่งใช้สัญลักษณ์ คือ σ^2

$$\sigma^2 = \frac{\sum(Xi-\mu)^2}{N}$$

สำหรับค่าความแปรปรวนของข้อมูลกลุ่มตัวอย่าง n

$$s^2 = \frac{\sum(Xi-X)^2}{n-1}$$

*ในบางกรณี ผลจากการคำนวณที่ได้อาจถูกนำเสนอในรูปแบบหนึ่ง ซึ่งเรียกว่าส่วนเบี่ยงเบนมาตรฐาน (*standard deviation*) ของข้อมูล โดยสามารถมีสูตรคำนวณดังนี้*

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum(Xi-\mu)^2}{N}} \quad ; \sum_{i-1}^{n}$$

หรือ

$$S = \sqrt{s^2} = \sqrt{\frac{\sum(Xi-X)^2}{n-1}} \quad ; \sum_{i-1}^{n}$$

ตัวอย่างที่ 3.9 *จากข้อมูลจำนวนรอยตำหนิในม้วนผ้า จำนวน 8 ข้อมูล ให้คำนวณหาค่าความแปรปรวนของข้อมูล*

วิธีทำ จากข้อมูลที่ได้มานำมาสร้างเป็นตารางสำหรับคำนวณหาค่าต่าง ๆ ได้ดังนี้

	Xi	*Xi-X*	$(Xi - X)^2$
	6	-2	*4*
	7	-1	*1*
	11	3	*9*
	7	-1	*1*
	8	0	*0*
	4	-4	*16*
	9	1	*1*
	12	4	*16*
ผลรวม	*64*	*0*	*48*

$$X = \frac{\mathbf{6+7+11+7+8+4+9+12}}{\mathbf{8}} = 8$$

วิธีที่ 1 การคำนวณค่าความแปรปรวนของตัวอย่าง โดยใช้สูตรที่กำหนดได้ดังนี้

$$s^2 = \frac{\sum(Xi-X)^2}{n-1} = \quad s^2 = \frac{(6-8)^2+(7-)^2+(11-8)^2+(7-8)^2+(8-8)^2+(4-8)^2+(9-8)^2+(12-8)^2+}{n-1}$$

s^2 = 48/7 = 6.86

และจะได้ค่าส่วนเบี่ยงเบนมาตรฐาน

$$s = \sqrt{6.86} = 2.62$$

วิธีที่ 2 การคำนวณค่าความแปรปรวนโดยใช้สูตรซึ่งจัดรูปแบบใหม่จากสูตรที่ใช้ในวิธีการที่ 1 คำนวณได้ดังนี้

	Xi	Xi^2
	6	36
	7	49
	11	121
	7	49
	8	64
	4	16
	9	81
	12	144
ผลรวม	64	560

สามารถคำนวณหาค่าความแปรปรวน ได้ดังนี้

$$S^2 = \frac{\sum_{i=1}^{n} Xi^2 - \frac{\sum_{i=1}^{n} Xi^2}{n}}{n-1} \quad = = \frac{560 - \frac{64^2}{8}}{8-1}$$

$$= \frac{560-512}{8-1} \quad = 6.86$$

และจะได้ ส่วนเบี่ยงเบนมาตรฐาน คือ

$$S = \sqrt{s^2} = \sqrt{6.86} = 2.62$$

จาก ผลการคำนวณด้วยวิธีทั้ง 2 สรุปได้ว่า ความแปรปรวนของข้อมูลนี้มี (S^2) ค่าเท่ากับ **6.86** และส่วนเบี่ยงเบนมาตรฐาน *(S)* มีค่า**เท่ากับ 2.62**
