

บทที่ 3

วิธีดำเนินงานวิจัย

การวิจัยครั้งนี้เป็นการเปรียบเทียบประสิทธิภาพวิธีประมาณค่าเฉลี่ยที่ดัดแปลงมาจากตัวประมาณ Horvitz-Thompson ภายใต้แผนการสุ่มตัวอย่างแบบกลุ่มปรับด้วยวิธีการประมาณค่าอย่างง่าย วิธีการประมาณค่าแบบอัตราส่วน และวิธีการประมาณค่าแบบอัตราส่วนที่ปรับปรุงโดยวิธีการ Rao-Blackwell เมื่อสุ่มหน่วยตัวอย่างอย่างง่าย โดยใช้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยเป็นเกณฑ์ในการเปรียบเทียบเพื่อหาวิธีประมาณค่าที่ดีที่สุดในแต่ละสถานการณ์ ข้อมูลที่ใช้ในการวิจัยได้จากการจำลองข้อมูลให้มีลักษณะตามสถานการณ์ที่กำหนดโดยใช้โปรแกรม Minitab 14 และการประมาณค่าพารามิเตอร์ใช้โปรแกรม MATLAB ซึ่งมีการวางแผน และกำหนดขั้นตอนดำเนินการวิจัยดังนี้

1. การสร้างข้อมูลในการวิจัย
2. ขั้นตอนการเลือกตัวอย่าง
3. การประมาณค่าเฉลี่ยภายใต้แผนการสุ่มตัวอย่างแบบกลุ่มปรับ
4. เกณฑ์การตัดสินใจ
5. แผนผังขั้นตอนดำเนินงานวิจัย

โดยมีรายละเอียดดังนี้

การสร้างข้อมูลในการวิจัย

ในการสร้างข้อมูลที่ใช้ในการวิจัยจะแบ่งออกเป็น 2 สถานการณ์หลักดังที่กล่าวไว้ในขอบเขตการวิจัยโดยในแต่ละสถานการณ์สร้างประชากรให้มีลักษณะที่แตกต่างกัน ภายใน 100 พื้นที่ย่อย

ตัวแปร Y คือ ตัวแปรที่ต้องการศึกษา เป็นข้อมูลเชิงปริมาณ

ตัวแปร X คือ ตัวแปรช่วยของสิ่งที่ต้องการศึกษา เป็นข้อมูลเชิงปริมาณในพื้นที่ย่อย

และในการสร้างข้อมูลเชิงปริมาณของ X และ Y จะต้องพิจารณาเงื่อนไขคือกำหนดให้ระดับความสัมพันธ์ระหว่างข้อมูล X และ Y ซึ่งแบ่งออกเป็น 3 ระดับดังนี้

ความสัมพันธ์ระดับต่ำ อยู่ในช่วง (0 , 0.4)

ความสัมพันธ์ระดับปานกลาง อยู่ในช่วง [0.4, 0.7)

ความสัมพันธ์ระดับสูง อยู่ในช่วง [0.7, 1.00)

ในการตรวจสอบระดับความสัมพันธ์ดังกล่าวจะใช้วิธีการหาค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างสองตัวแปร ถ้าข้อมูลที่จำลองขึ้นมีความสัมพันธ์ไม่ตรงตามที่กำหนดไว้ ก็จะทำการศึกษาข้อมูลจนกระทั่งเป็นไปตามข้อตกลงทุกประการ ส่วนประชากรที่ใช้ในการวิจัยมีหลักการสร้างดังนี้

หลักการสร้างประชากรที่มีลักษณะที่หายาก

กระบวนการปัวซองคลัสเตอร์ หรือเรียกอีกอย่างหนึ่งว่า Centre – Satellite Process เป็นกระบวนการจำลองสถานการณ์โดยอาศัยระยะห่างของค่าสังเกตแต่ละตัว มาสร้างสถานการณ์หรือจำลองประชากรให้มีลักษณะเกาะกลุ่มกันภายในพื้นที่ที่กำหนด พัฒนาการครั้งแรกโดย Neyman และ Scott (1985) ซึ่งต่อมา A.D.Cliff และ J.K.Ord ได้สรุปกระบวนการปัวซองคลัสเตอร์โดยการจำลองด้วยเครื่องคอมพิวเตอร์ซึ่งทำอย่างต่อเนื่องโดยมีหลักการที่สำคัญ 3 ขั้นตอนคือ

1. หาจำนวนจุดหลัก สร้างด้วยกระบวนการปัวซอง ที่มีความหนาแน่น λ ต่อหนึ่งหน่วยพื้นที่ มี 2 ขั้นตอนคือ

1.1 หาตำแหน่งของเหตุการณ์ ซึ่งจำลองจากตัวแปรสุ่มอิสระที่มีฟังก์ชันความหนาแน่นแบบ Exponential (λ) และจำลองจากโปรแกรม Minitab 14 ด้วย $\lambda = 30$ ในช่วงพื้นที่ 100 หน่วย และจำลองประชากรออกเป็น 2 สถานการณ์หลักซึ่งแต่ละสถานการณ์มีลักษณะของสิ่งที่สนใจกระจายอยู่ในรูปแบบที่แตกต่างกัน

1.2 นับจำนวนเหตุการณ์ที่เกิดในช่วง (0,100) โดยเอาตำแหน่งที่ไม่เกิน 100 เท่านั้น

2. หาจำนวนบริวาร ของแต่ละจุดหลักซึ่งเป็นค่าสังเกตที่เป็นอิสระต่อกันมาจากการแจกแจงความน่าจะเป็นที่กำหนด โดยงานวิจัยนี้ใช้การแจกแจงแบบปัวซอง และจำลองจากโปรแกรม Minitab 14 ด้วย $\lambda = 30$

3. หาที่ตั้งของบริวาร มีความสัมพันธ์กับตำแหน่งของจุดหลักซึ่งเป็นอิสระต่อกัน มีการแจกแจงแบบปกติ 2 ตัวแปร

ในการดำเนินการสร้างประชากรที่มีลักษณะที่หายาก ผู้วิจัยขอแนะนำเสนอรายละเอียดการสร้างประชากรเฉพาะกรณีหน่วยขอบแยกกัน 2 กลุ่มเท่านั้น ส่วนประชากรกรณีอื่น ๆ สามารถพิจารณาได้ในลักษณะเดียวกัน

ขั้นตอนการจำลองประชากรที่มีลักษณะหายากมีรายละเอียดดังนี้

1. หาจำนวนจุดหลัก สร้างด้วยกระบวนการปัวซอง ที่มีความหนาแน่น λ ต่อหนึ่งหน่วยพื้นที่ (Unit Area) มี 2 ขั้นตอนคือ

1.1 หาตำแหน่งของเหตุการณ์ซึ่งจำลองจากตัวแปรสุ่มอิสระที่มีฟังก์ชันความหนาแน่นแบบ Exponential (λ) และจำลองจากโปรแกรม Minitab 14 ด้วย $\lambda = 30$ ได้ผลลัพธ์ดังนี้

MTB > Random 5 c1

SUBC > Exponential 30

MTB > print c1

C1

38.2631 35.0943 39.2180 82.6139 65.6530

1.2 นับจำนวนเหตุการณ์ที่เกิดในช่วง (0,100) โดยเอาตำแหน่งที่ไม่เกิน 100 เท่านั้น ผลลัพธ์ที่ได้คือมีจำนวนจุดหลัก 2 จุด คือ (38.2631 + 35.0943 = 73.3574) ซึ่งไม่รวมจุดที่ 3 ที่มีค่า 39.2180 ซึ่งเกิน 100

2. หาจำนวนบริวาร ของแต่ละจุดหลักซึ่งเป็นค่าสังเกตที่เป็นอิสระต่อกันมาจากการแจกแจงความน่าจะเป็นที่กำหนด โดยงานวิจัยนี้ใช้การแจกแจงแบบปัวซอง และจำลองจากโปรแกรม Minitab 14 ด้วย $\lambda = 30$ ได้ผลลัพธ์ดังนี้

MTB > Random 5 c1

SUBC > Exponential 30

MTB > print c1

C1

24 28

ดังนั้น

กลุ่มที่ 1 มีจำนวนบริวาร 24 ตัว

กลุ่มที่ 2 มีจำนวนบริวาร 28 ตัว

3. หาที่ตั้งของบริวาร มีความสัมพันธ์กับตำแหน่งของจุดหลักซึ่งเป็นอิสระต่อกัน มีการแจกแจงแบบปกติ 2 ตัวแปร จำลองจากโปรแกรม Minitab 14 ได้ผลลัพธ์ดังนี้

3.1 Choose Data > Copy > Columns to Matrix.

3.2 In Copy from columns, enter C2-C3.

3.3 In Store Copied Data, under In current worksheet, in matrix, enter M1.

Click OK.

3.4 Choose Calc > Random Data > Multivariate Normal.

3.5 In Generate _ rows of data, enter 24 และ 28

3.6 In Store in column(s): enter C7-C8 และ C10-C11

3.7 In Mean column: enter C1.

3.8 In Variance-Covariance Matrix: enter M1. Click OK.

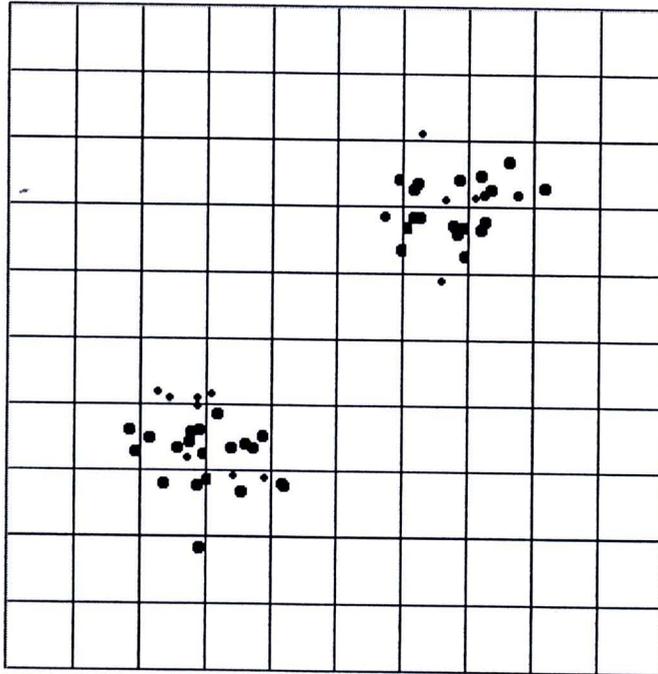
จุดหลักที่ 1 มีจำนวน 24 ตัว มีตำแหน่งดังนี้

ตัวแปร x	ตัวแปร y	ตัวแปร x	ตัวแปร y
-0.13216	-0.54078	-0.6658	-0.4388
-0.55464	0.25124	1.75325	1.6368
0.58845	0.4153	0.44423	-0.07828
-0.84887	-0.46815	0.26691	0.66925
-0.38235	-0.01228	-0.64854	-1.13505
0.27358	0.21053	-1.4776	-1.32582
-0.71009	-0.60903	1.54749	0.95681
0.03009	-0.64639	-0.28705	0.15097
-0.06201	-1.06288	0.07636	-0.04871
-0.89496	-0.70492	0.07669	0.25502
-1.06091	-1.24888	-1.94968	-1.98580
0.92397	1.40132	-0.17406	-0.60014

จุดหลักที่ 2 มีจำนวน 28 ตัว มีตำแหน่งดังนี้

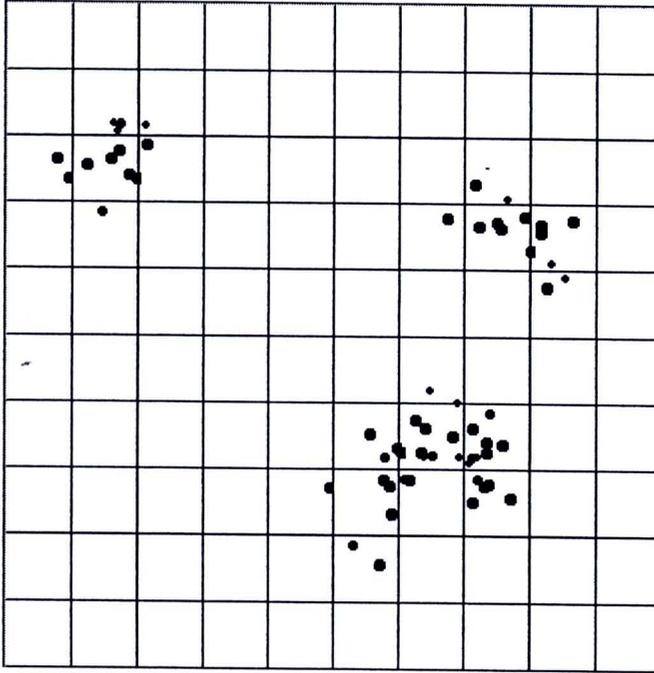
ตัวแปร x	ตัวแปร y	ตัวแปร x	ตัวแปร y
-0.68739	-0.60039	-0.86609	-0.99349
1.52703	-0.02338	-0.07764	-0.94985
0.47187	1.3179	-0.85202	-1.66508
-1.42746	-0.87319	0.83477	1.07267
-0.30205	0.36178	-0.28761	-0.59992
-0.6064	0.33108	0.75047	1.05844
-0.39166	-0.85091	0.90159	0.58562
-2.01651	-1.94199	0.07714	0.50855
0.00719	-0.11800	-0.32716	0.69242
0.73885	1.34408	0.98713	0.13160
-0.88252	-1.34563	-0.29595	0.19970
-0.05057	-0.17976	-0.44148	-1.13529
-0.87833	-1.14517	0.17933	-0.10763
-0.55735	-1.73831	1.62260	1.25489

กรณีประชากรสถานการณ์ที่ 1 หน่วยขอบแยกกัน แสดงดังแผนภาพ 11, 12 และ 13 ตามลำดับ โดยสามารถพิจารณารายละเอียดได้ดังนี้



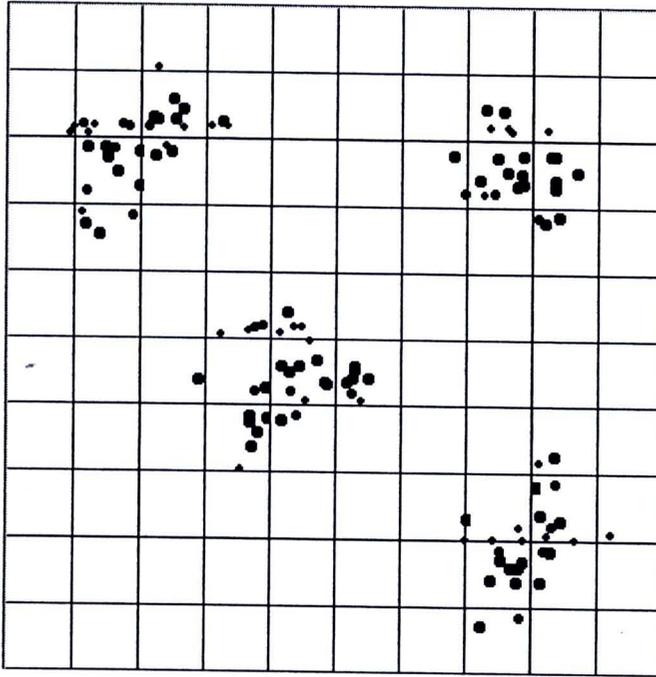
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	1	0	0	0
0	0	0	0	0	1	4	6	1	0
0	0	0	0	0	3	5	2	0	0
0	0	0	0	0	0	1	0	0	0
0	0	4	1	0	0	0	0	0	0
0	2	7	5	0	0	0	0	0	0
0	0	3	3	2	0	0	0	0	0
0	0	1	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0

ภาพ 11 แสดงที่ตั้ง และจำนวนของประชากรสถานการณ์ที่ 1 กรณีหน่วยขอบแยกกัน 2 กลุ่ม



0	0	0	0	0	0	0	0	0	0
0	3	1	0	0	0	0	0	0	0
2	5	1	0	0	0	0	2	0	0
0	1	0	0	0	0	1	5	4	0
0	0	0	0	0	0	0	0	2	0
0	0	0	0	0	0	2	0	0	0
0	0	0	0	0	4	7	8	0	0
0	0	0	0	1	3	2	5	0	0
0	0	0	0	0	2	0	0	0	0
0	0	0	0	0	0	0	0	0	0

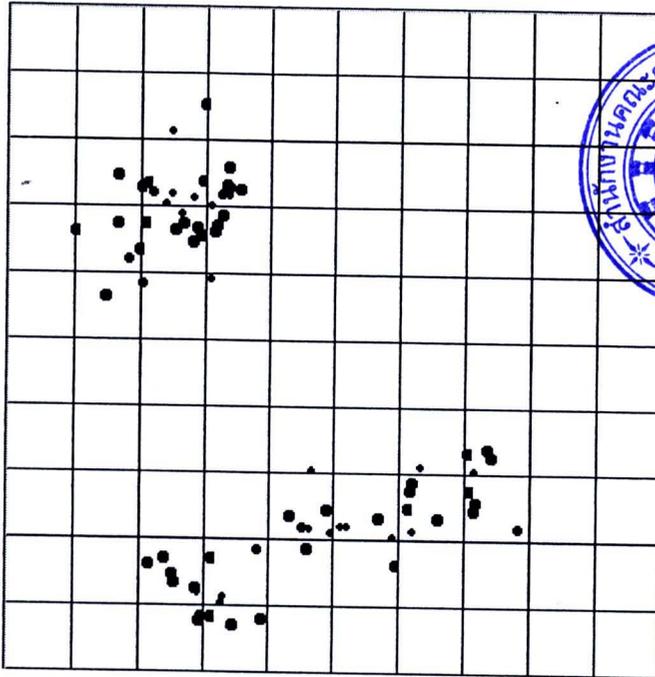
ภาพ 12 แสดงที่ตั้ง และจำนวนของประชากรสถานการณ์ที่ 1
กรณีหน่วยขอบแยกกัน 3 กลุ่ม



0	0	1	0	0	0	0	0	0	0
2	5	7	3	0	0	0	5	1	0
0	8	3	0	0	0	2	9	5	0
0	4	0	0	0	0	0	0	3	0
0	0	0	4	5	0	0	0	0	0
0	0	1	2	8	7	0	0	0	0
0	0	0	6	2	0	0	0	2	0
0	0	0	0	0	0	2	4	6	1
0	0	0	0	0	0	0	7	3	0
0	0	0	0	0	0	0	2	0	0

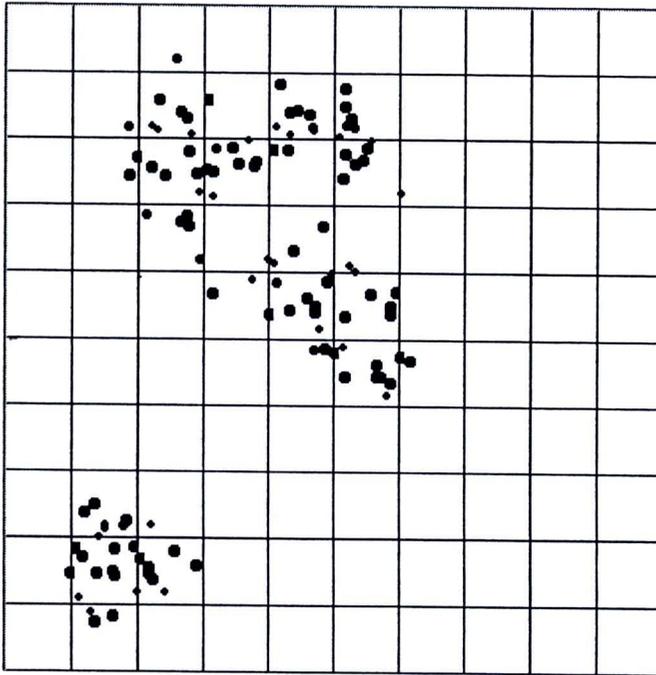
ภาพ 13 แสดงที่ตั้ง และจำนวนของประชากรสถานการณ์ที่ 1
กรณีหน่วยขอบแยกกัน 4 กลุ่ม

กรณีประชากรสถานการณ์ที่ 2 หน่วยขอพร้อมกัน แสดงดังแผนภาพ 14, 15 และ 16 ตามลำดับ โดยสามารถพิจารณารายละเอียดได้ดังนี้



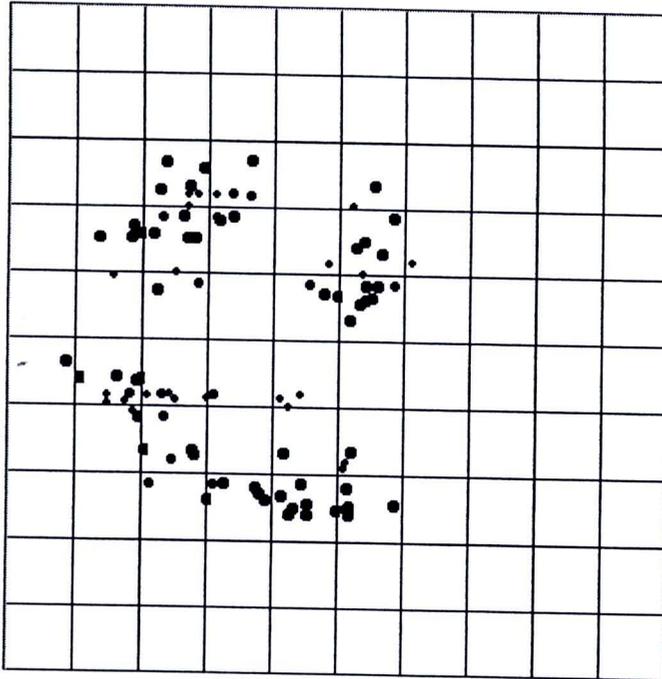
0	0	0	0	0	0	0	0	0	0
0	0	2	0	0	0	0	0	0	0
0	3	5	7	0	0	0	0	0	0
1	4	6	3	0	0	0	0	0	0
0	2	0	1	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	1	0	2	3	0	0
0	0	0	0*	5	4	6	3	0	0
0	0	7	3	1*	1	0	0	0	0
0	0	3	2	0	0	0	0	0	0

ภาพ 14 แสดงที่ตั้ง และจำนวนของประชากรสถานการณ์ที่ 2 กรณีหน่วยขอพร้อมกัน 2 กลุ่ม



0	0	1	0	0	0	0	0	0	0
0	1	7*	8*	8	7	0	0	0	0
0	2	6	8*	8*	6	0	0	0	0
0	0	5	8*	4	2*	0	0	0	0
0	0	0*	3	7	5	0	0	0	0
0	0	0	0	3	8	1	0	0	0
0	0	0	0	0	0	0	0	0	0
0	7	1	0	0	0	0	0	0	0
2	9	6	0	0	0	0	0	0	0
0	3	0	0	0	0	0	0	0	0

ภาพ 15 แสดงที่ตั้ง และจำนวนของประชากรสถานการณ์ที่ 2
กรณีหน่วยขอร่วมกัน 3 กลุ่ม



0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	7	4	0	2	0	0	0	0
0	5	6	3	*	5	1	0	0	0
0	*	*	*	3	7	0	0	0	0
2	7	5	*	3	*	0	0	0	0
0	3	4	*	*	3	0	0	0	0
0	0	*	5	7	5	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0

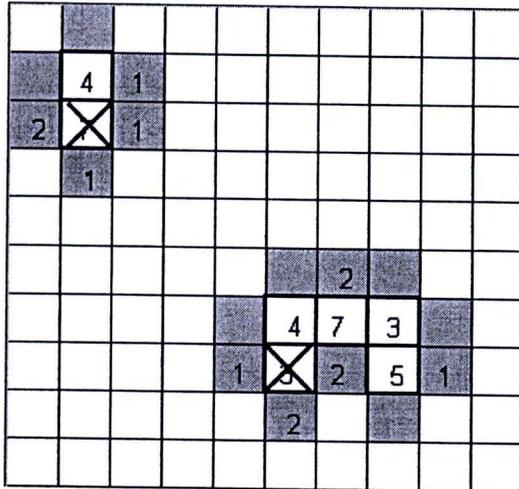
ภาพ 16 แสดงที่ตั้ง และจำนวนของประชากรสถานการณที่ 2
กรณีหน่วยขอบรวมกัน 4 กลุ่ม

ขั้นตอนการเลือกตัวอย่าง

กำหนดขนาดพื้นที่ย่อยที่ใช้ในการวิจัย คือ 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 16, 20 และ 24 หน่วย และเลือกหน่วยตัวอย่างภายใต้แผนการสุ่มตัวอย่างแบบกลุ่มปรับดังนี้

1. เลือกหน่วยตัวอย่างด้วยวิธีการสุ่มตัวอย่างแบบง่ายไม่แทนที่จากประชากรที่มีลักษณะที่หายาก 2 สถานการณ์หลัก
2. ถ้าพบว่าหน่วยตัวอย่างที่ได้จากการสุ่มเป็นไปตามเงื่อนไขคือ $C = \{y : y \geq 3\}$ โดยที่ C เป็น set ของค่าที่มากกว่าหรือเท่ากับ 3 โดยถ้า y_i ที่อยู่ใน C มีค่ามากกว่าหรือเท่ากับ 3 ที่เป็นไปตามเงื่อนไขที่กำหนดก็จะพิจารณาหน่วยบริเวณใกล้เคียงเข้าไปเป็นหน่วยตัวอย่าง

โดยรูปแบบ และวิธีการเลือกตัวอย่างแสดงรายละเอียดดังภาพ 17 ซึ่งเป็นการสุ่มพื้นที่ย่อยด้วยวิธีการสุ่มตัวอย่างแบบง่าย จากนั้นก็จะพิจารณาพื้นที่ย่อยบริเวณใกล้เคียง โดยที่เงื่อนไขของการรวมเป็นเครือข่ายคือ $y_i \geq 3$ ถ้าเป็นไปตามเงื่อนไขก็จะถูกรวมเป็นเครือข่าย แต่ถ้าไม่เป็นไปตามเงื่อนไขก็จะไม่พิจารณา และส่วนที่อยู่รอบเครือข่ายที่เป็นสี่เหลี่ยมจะเรียกว่าหน่วยขอบ



ภาพ 17 แสดงการเลือกตัวอย่าง และการขยายพื้นที่ตามเงื่อนไขที่กำหนด

การประมาณค่าเฉลี่ยภายใต้แผนการสุ่มตัวอย่างแบบกลุ่มปรับ

1. วิธีประมาณค่าอย่างง่าย เป็นการประมาณค่าเฉลี่ยของตัวแปรที่สนใจศึกษา ที่ตัวประมาณมีการดัดแปลงให้เหมาะสมกับแผนการสุ่มตัวอย่างแบบกลุ่มปรับ
2. วิธีการประมาณค่าแบบอัตราส่วน เป็นวิธีการประมาณค่าเฉลี่ยของตัวแปรที่สนใจศึกษา และมีการใช้ประโยชน์จากตัวแปรช่วยมาพิจารณาในการประมาณค่า

3. วิธีการประมาณค่าแบบอัตราส่วนที่ปรับปรุงโดยวิธีการ Rao-Blackwell เป็นวิธีการประมาณค่าที่ใช้ประโยชน์จากตัวแปรช่วย และใช้วิธีการ Rao-Blackwell มาปรับปรุงวิธีการเพื่อทำให้ค่าประมาณที่ได้เป็นสถิติพอเพียงต่ำสุด และมีความแปรปรวนต่ำ ซึ่งเป็นคุณสมบัติที่ดีทางสถิติ

เกณฑ์การตัดสินใจ

เกณฑ์ที่ใช้ในการประเมินว่าวิธีการประมาณค่าในรูปแบบการสุ่มตัวอย่างแบบกลุ่มปรับวิธีการใดที่มีประสิทธิภาพมากที่สุดจะพิจารณาจากค่าความคลาดเคลื่อนกำลังสองเฉลี่ยโดยวิธีที่ให้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุดจะเป็นวิธีการประมาณค่าที่มีประสิทธิภาพมากที่สุด โดยมีวิธีการคำนวณดังนี้

$$Bias(\bar{y}) = \frac{\sum_{t=1}^{1000} Bias(\bar{y}_t)}{1000} \quad \text{โดยที่ } Bias(\bar{y}_t) = \bar{y}_t - \mu$$

เมื่อ

\bar{y}_t คือ ค่าประมาณค่าเฉลี่ยรอบที่ t ของแต่ละวิธี
 t คือ จำนวนรอบของการทำซ้ำ

โดยที่

$Bias(\bar{y})$ คือ ค่าประมาณค่าความเอนเอียงของค่าประมาณค่าเฉลี่ย
 $Bias(\bar{y}_t)$ คือ ค่าประมาณค่าความเอนเอียงของค่าประมาณค่าเฉลี่ยรอบที่ t

ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย

$$MSE(\bar{y}) = \frac{\sum_{i=1}^{1000} MSE(\bar{y}_i)}{1000} \quad \text{โดยที่ } MSE(\bar{y}_t) = Var(\bar{y}_t) + Bias(\bar{y}_t)^2$$

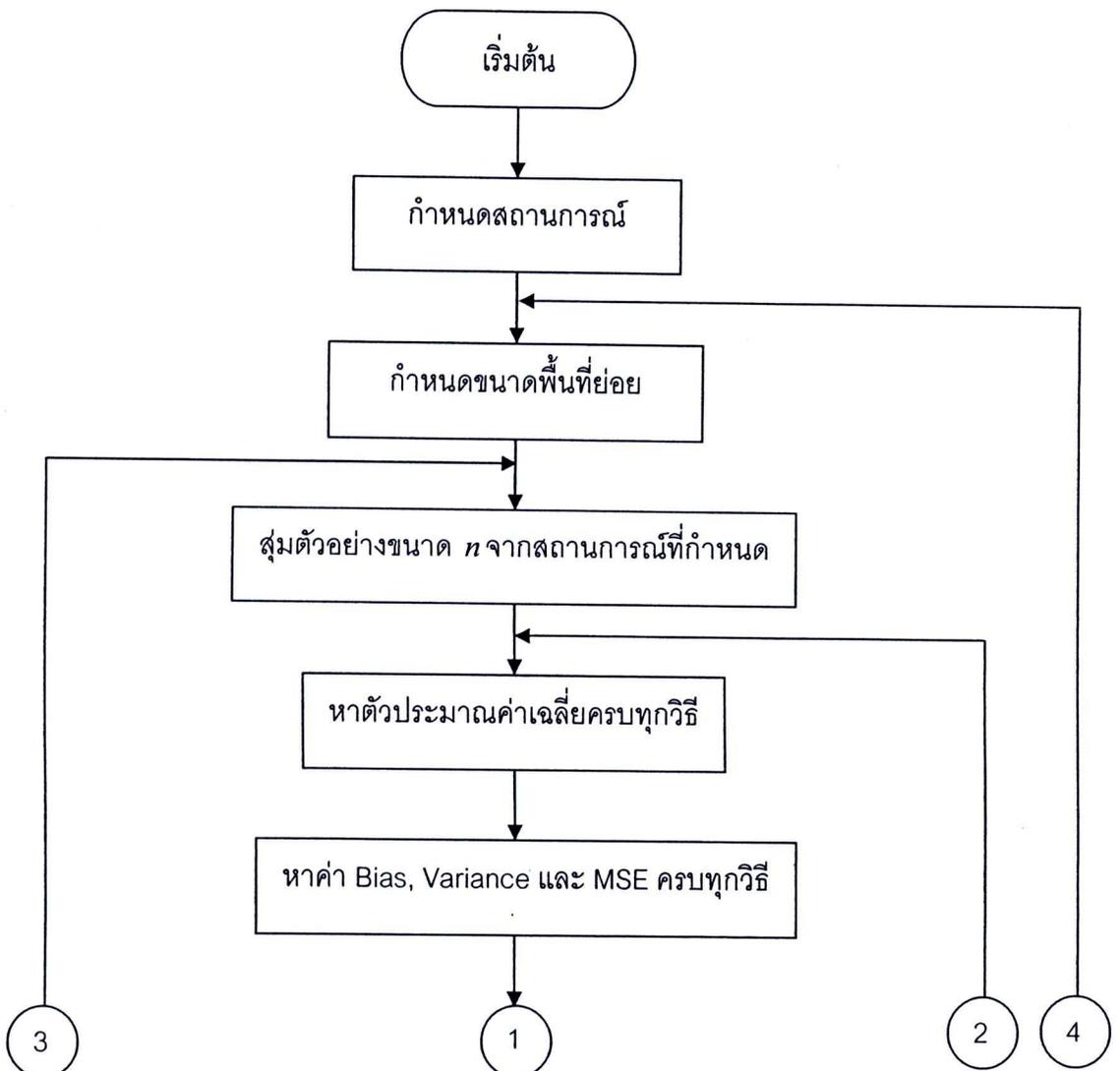
โดยที่

$MSE(\bar{y})$ คือ ค่าประมาณความคลาดเคลื่อนกำลังสองของค่าประมาณค่าเฉลี่ย
 $MSE(\bar{y}_t)$ คือ ค่าความคลาดเคลื่อนกำลังสองของค่าเฉลี่ยรอบที่ t

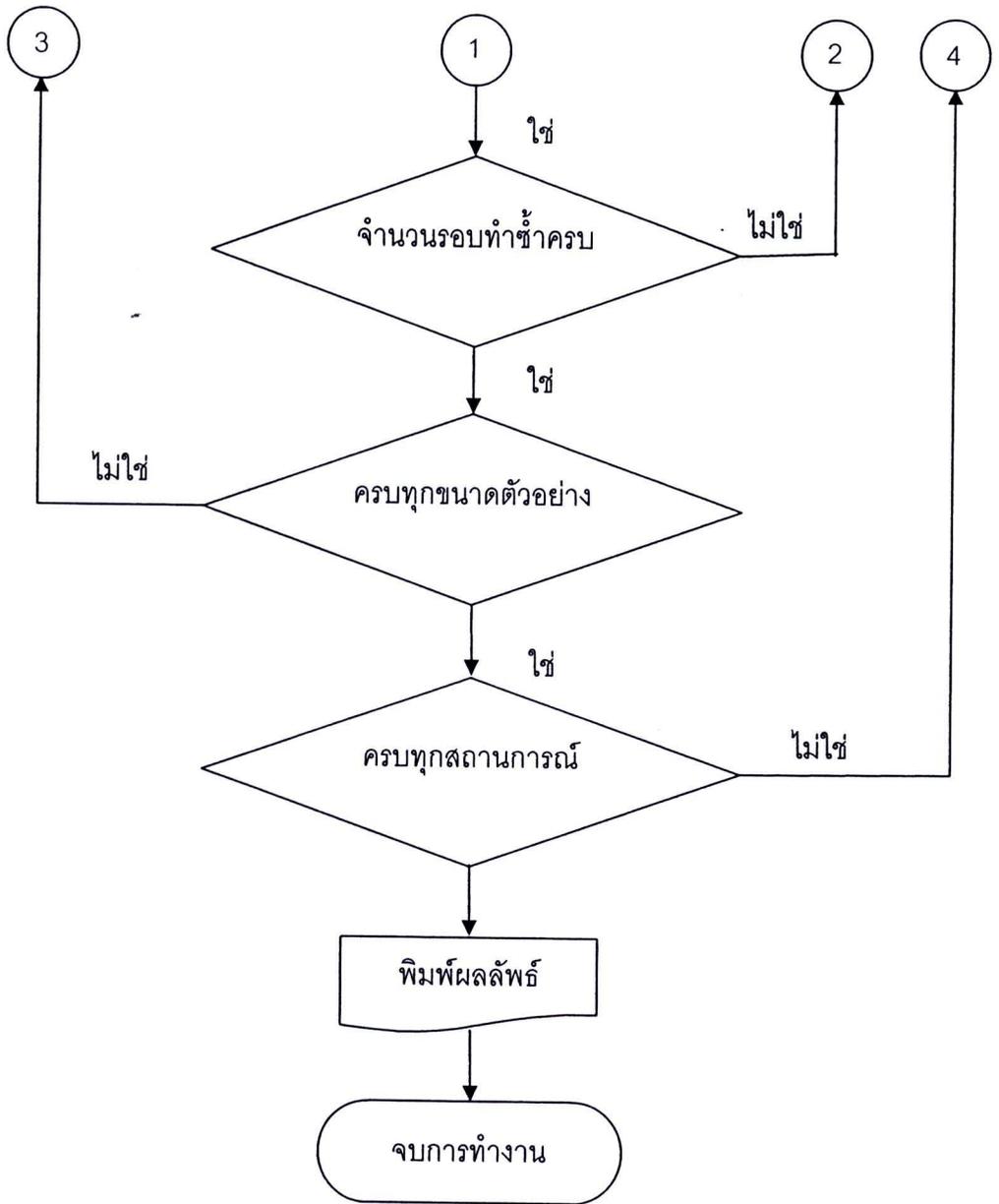
$Var(\bar{y}_t)$ คือ ค่าประมาณความแปรปรวนของค่าประมาณค่าเฉลี่ยรอบที่ t ของแต่ละวิธีที่ได้กล่าวไว้ในทฤษฎีบทที่ 2

แผนผังขั้นตอนดำเนินงานวิจัย

ในการวิจัยครั้งนี้ ผู้วิจัยได้ใช้โปรแกรม MATLAB ในการสร้าง และวิเคราะห์ข้อมูลต่าง ๆ ซึ่งสามารถแสดงแผนผังขั้นตอนดำเนินงานวิจัยได้ดังนี้



ภาพ 18 แสดงแผนผังแสดงลำดับการทำงานหลักของโปรแกรม



ภาพ 18 (ต่อ)