บทที่ 3

วิธีดำเนินการวิจัย

ในการวิจัยครั้งนี้ มีจุดประสงค์ที่จะทำการเปรียบเทียบวิธีการประมาณขนาดประชากร เมื่อมีการเลือกตัวอย่างแบบแคบเจอร์ - รีแคบเจอร์ โดยวิธีการประมาณ 3 วิธี คือ วิธีการประมาณโดยอาศัยตัวประมาณ Petersen วิธีการประมาณโดยอาศัย Gibbs Sampler และวิธีการประมาณโดยอาศัยตัวประมาณที่พัฒนาจากตัวประมาณ Petersen โดยมีขั้นตอนในการวิจัยดังนี้

1. กำหนดค่าขนาดประชากรที่ต้องการทำการประมาณ กำหนดขนาดประชากรดังนี้
   - ประชากรที่มีขนาดเล็ก คือ N = 1000 , 3000 , 5000 , 7000
   - ประชากรที่มีขนาดปานกลาง คือ N = 10000 , 30000 , 50000 , 70000

2. คำนวณค่าประมาณจากวิธีการประมาณทั้งสาม โดยมีรายละเอียดดังนี้

2.1 วิธีประมาณโดยอาศัยตัวประมาณ Petersen จะทำการประมาณโดยแบ่งออกเป็น 2 กรณี คือ

กรณีที่ 1 : เมื่อจำนวนตัวอย่างที่ถูกเลือกซ้ำในการเลือกครั้งที่ 2 (m_2) ถูกต้องตามทฤษฎีความน่าจะเป็น คือ ถ้าขนาดประชากรเท่ากับ N และขนาดตัวอย่างที่ทำการเลือกในครั้งแรกมีค่าเท่ากับ n_1 และขนาดตัวอย่างที่ทำการเลือกในครั้งที่สองมีค่าเท่ากับ n_2 ดังนั้น m_2 ควรจะมีค่าเท่ากับสมการต่อไปนี้

$$m_2 = \frac{n_1 n_2}{N} \tag{2.1.1}$$

กรณีที่ 2 : เมื่อจำนวนตัวอย่างที่ถูกเลือกซ้ำ (m_2) ไม่เป็นไปตามทฤษฎีความน่าจะเป็น คือ ถ้าขนาดประชากรเท่ากับ N และขนาดตัวอย่างที่ทำการเลือกในครั้งแรกมีค่าเท่ากับ n_1 ขนาดตัวอย่างที่ทำการเลือกในครั้งที่สองมีค่าเท่ากับ n_2 และจำนวนตัวอย่างที่ถูกเลือกซ้ำผิดพลาดไปร้อยละ q_1 ดังนั้น m_2 ควรจะมีค่าเท่ากับสมการต่อไปนี้

$$m_2 = \frac{n_1 n_2 q_1}{N \times 100} \tag{2.1.2}$$

จากทั้งสองกรณีทำการคำนวณค่าประมาณโดยวิธี Petersen ตามสมการต่อไปนี้

$$N_P = \frac{(n_1+1)(n_2+1)}{(m_2+1)} - 1 \tag{2.1.3}$$

คำนวณค่าเฉลี่ยร้อยละของความคลาดเคลื่อนสัมพัทธ์สัมบูรณ์ดังนี้

$$MAPE = \frac{1}{I}\sum_{i=1}^{K}\left|\frac{(N_i - N_{Pi})}{N_i}\right| \times 100$$

(2.1.4)

2.2 วิธีประมาณโดยอาศัย Gibbs Sampler จะทำการประมาณ โดยแบ่งออกเป็น 4 กรณี คือ

กรณีที่ 1 : จำนวนตัวอย่างที่ถูกเลือกซ้ำในการเลือกตัวอย่างครั้งที่ 2 (m_2) และ 3 (m_3) ถูกต้องตามทฤษฎีความน่าจะเป็น เป็นไปตามทฤษฎีความน่าจะเป็น ดังนั้นค่า m_2 และ m_3 จะคำนวณได้จากสมการต่อไปนี้

$$m_2 = \frac{n_1 n_2}{N} \tag{2.2.1}$$

$$m_3 = \frac{(n_1 + n_2 - m_2)n_3}{N} \tag{2.2.2}$$

กรณีที่ 2 : จำนวนตัวอย่างที่ถูกเลือกซ้ำในการเลือกตัวอย่างครั้งที่ 2 (m_2) ผิดพลาดร้อยละ q_1 แต่จำนวนตัวอย่างที่ถูกเลือกซ้ำในการเลือกตัวอย่างครั้งที่ 3 (m_3) ถูกต้องตามทฤษฎี จากกรณีนี้จะได้ว่าค่า m_2 สามารถคำนวณได้ดังนี้

$$m_2 = \frac{n_1 n_2 q_1}{N \times 100} \tag{2..2.3}$$

ส่วนค่า m_3 จะทำการคำนวณเช่นเดียวกับกรณีที่ 1

กรณีที่ 3 : จำนวนตัวอย่างที่ถูกเลือกซ้ำในการเลือกตัวอย่างครั้งที่ 3 (m_3) ผิดพลาดร้อยละ q_2 แต่จำนวนตัวอย่างที่ถูกเลือกซ้ำในการเลือกตัวอย่างครั้งที่ 2 (m_2) ถูกต้องตามทฤษฎี จากกรณีนี้จะได้ว่าค่า m_3 สามารถคำนวณได้ดังนี้

$$m_3 = \frac{(n_1 + n_2 - m_2)n_3 q_2}{N \times 100} \tag{2.2.4}$$

ส่วนค่า m_2 จะทำการคำนวณเช่นเดียวกับกรณีที่ 1

กรณีที่ 4 : จำนวนตัวอย่างที่ถูกเลือกซ้ำในการเลือกตัวอย่างครั้งที่ 2 (m_2) และจำนวนตัวอย่างที่ถูกเลือกซ้ำในการเลือกตัวอย่างครั้งที่ 3 (m_3) ผิดพลาดร้อยละ q_1 และ q_2 ตามลำดับ ในกรณีนี้จะคำนวณค่า m_2 , m_3 ได้จากสมการ (2.2.3) และ (2.2.4) ตามลำดับ

การคำนวณสำหรับตัวประมาณที่อาศัย Gibbs Sampler มีขั้นตอนดังต่อไปนี้

ขั้นตอนที่ 1 : กำหนดค่า N เริ่มต้น, n_1, n_2, n_3, a และ b

ขั้นตอนที่ 2 : คำนวณค่า m_1, m_2, m_{23} และ r ตามสมการของแต่ละกรณี

ขั้นตอนที่ 3 : จำลองค่า p_i ; i = 1, 2, 3 ให้มีการแจกแจงแบบเบตา (Beta Distribution) ที่มีพารามิเตอร์เป็น (n_i+a,N-n_i+b) ซึ่งสามารถจำลองได้ดังนี้

3.1 ทำการจำลอง Y1 และ Y2 ให้มีการแจกแจงแบบแกมมา ที่มีพารามิเตอร์เป็น (n_i+a,1) และ (N-n_i+b) ตามลำดับ โดยที่ Y1 และ Y2 เป็นอิสระต่อกัน

3.2 ค่า p_i ที่ต้องการจำลองจะมีค่าเท่ากับ Y1/(Y1+Y2)

ขั้นตอนที่ 4 : จำลองค่า N ให้มีการแจกแจงแบบทวินามปกติเชิงนิเสธ(Negative Binomial) โดยที่มีพารามิเตอร์เป็น $(r - 1, 1 - \prod_{i=1}^{3}(1 - p_i))$ ซึ่งสามารถจำลองได้ดังนี้

4.1 ทำการจำลอง Xi ให้มีการแจกแจงแบบเรขาคณิต (Geometric Distribution) ที่มีพารามิเตอร์เป็น $(1 - \prod_{i=1}^{3}(1 - p_i))$ เป็นจำนวน r ตัว โดยที่ X_i ต้องอิสระต่อกัน

4.2 ค่า N ที่ต้องการจำลองจะมีค่าเท่ากับผลรวมของ X_i

<u>หมายเหตุ</u> : วิธีการจำลองการแจกแจงแบบแกมมา และแบบเรขาคณิตดูได้ที่ภาคผนวก

ขั้นตอนที่ 5 : ทำการจำลองค่า $N^{(k)}$ และ $P^{(k)}$ เป็นจำนวน m รอบ แล้วนำค่าสุดท้ายไปเป็นค่าเริ่มต้นสำหรับการจำลองครั้งต่อ ๆ ไป โดยเก็บค่าที่ $N^{(m)}$ ไว้ ทำเช่นนี้ไปจนกระทั่งได้ค่า $N^{(m)}$ เป็นจำนวน K ตัว (ในการวิจัยครั้งนี้ m มีค่าเท่ากับ 100 และ K มีค่าเท่ากับ 20)

ขั้นตอนที่ 6 : ทำการประมาณขนาดประชากร โดยตัวประมาณที่อาศัย Gibbs Sampler จะมีค่าดังสมการ

$$N_g = \sum_{k=1}^{20} \frac{N_k^{(m)}}{20} \tag{2.2.5}$$

เมื่อคำนวณค่าประมาณได้แล้ว จะคำนวณค่าร้อยละของความคลาดเคลื่อนสัมพัทธ์สัมบูรณ์จากสมการต่อไปนี้

$$MAPE = \frac{\sum_{i=1}^{I} \left| \frac{(N_i - N_{gi})}{N_i} \right| \times 100}{I} \tag{2.2.6}$$

2.3 วิธีประมาณโดยตัวประมาณที่พัฒนามาจากตัวประมาณ Petersen จะทำการประมาณ โดยแบ่งออกเป็น 5 กรณี คือ

กรณีที่ 1 : จำนวนตัวอย่างที่ถูกเลือกซ้ำในการเลือกตัวอย่างครั้งที่ 2 (m_2) และ 3 (m_3) ถูกต้องตามทฤษฎีความน่าจะเป็น และจำนวนตัวอย่างที่ถูกเลือกซ้ำทั้งในการเลือกตัวอย่างครั้งที่ 2 และ 3 (m_{23}) เป็นไปตามทฤษฎีความน่าจะเป็น ดังนั้นค่า m_2 , m_3 , m_{23} จะคำนวณได้จากสมการต่อไปนี้

$$m_2 = \frac{n_1 n_2}{N} \tag{2.3.1}$$

$$m_3 = \frac{(n_1 + n_2 - m_2) n_3}{N} \tag{2.3.2}$$

$$m_{23} = \frac{m_2 n_3}{N} \tag{2.3.3}$$

กรณีที่ 2 : จำนวนตัวอย่างที่ถูกเลือกซ้ำในการเลือกตัวอย่างครั้งที่ 2 (m_2) ผิดพลาดร้อยละ q_1 แต่จำนวนตัวอย่างที่ถูกเลือกซ้ำในการเลือกตัวอย่างครั้งที่ 3 (m_3) และจำนวนตัวอย่างซ้ำที่ถูกเลือกทั้งในการเลือกครั้งที่ 2 และ 3 (m_{23}) ถูกต้องตามทฤษฎี จากกรณีนี้จะได้ว่า ค่า m_2 สามารถคำนวณได้ดังนี้

$$m_2 = \frac{n_1 n_2 q_1}{N \times 100} \tag{2.3.4}$$

ส่วนค่าอื่น ๆ จะทำการคำนวณเช่นเดียวกับกรณีที่ 1

กรณีที่ 3 : จำนวนตัวอย่างที่ถูกเลือกซ้ำในการเลือกตัวอย่างครั้งที่ 3 (m_3) ผิดพลาดร้อยละ q_2 แต่จำนวนตัวอย่างที่ถูกเลือกซ้ำในการเลือกตัวอย่างครั้งที่ 2 (m_2) และจำนวนตัวอย่างที่ถูกเลือกซ้ำทั้งในการเลือกครั้งที่ 2 และ 3 (m_{23}) ถูกต้องตามทฤษฎี จากกรณีนี้จะได้ว่า ค่า m_2 สามารถคำนวณได้ดังนี้

$$m_3 = \frac{(n_1 + n_2 - m_2) n_3 q_2}{N \times 100} \tag{2.3.5}$$

ส่วนค่าอื่น ๆ จะทำการคำนวณเช่นเดียวกับกรณีที่ 1

กรณีที่ 4 : จำนวนตัวอย่างที่ถูกเลือกซ้ำทั้งในการเลือกครั้งที่ 2 และ 3 (m_{23})ผิดพลาดร้อยละ q_3 แต่จำนวนตัวอย่างที่ถูกเลือกซ้ำในการเลือกตัวอย่างครั้งที่ 2 (m_2) และจำนวนตัวอย่างที่ถูกเลือกซ้ำในการเลือกตัวอย่างครั้งที่ 3 (m_3) ถูกต้องตามทฤษฎี จากกรณีนี้จะได้ว่า ค่า m_{23} สามารถคำนวณได้ดังนี้

$$m_{23} = \frac{m_2 n_3 q_3}{N \times 100} \tag{2.2.6}$$

ส่วนค่าอื่น ๆ จะทำการคำนวณเช่นเดียวกับกรณีที่ 1

กรณีที่ 5 : จำนวนตัวอย่างที่ถูกเลือกซ้ำในการเลือกตัวอย่างครั้งที่ 2 (m_2) และจำนวนตัวอย่างที่ถูกเลือกซ้ำในการเลือกตัวอย่างครั้งที่ 3 (m_3) ผิดพลาดร้อยละ q_1 และ q_2 ตาม

ลำดับ แต่จำนวนตัวอย่างที่ถูกเลือกซ้ำทั้งในการเลือกครั้งที่ 2 และ 3 (m_{23}) ถูกต้องตามทฤษฎีความน่าจะเป็น ในกรณีนี้จะคำนวณค่า m_2 , m_3 และ m_{23} ได้จากสมการ (2.3.4) , (2.3.5) และ (2.3.3) ตามลำดับ

การคำนวณค่าประมาณจากวิธีที่พัฒนาจากวิธี Petersen นั้น มีขั้นตอนในการคำนวณดังนี้

ขั้นตอนที่ 1. : คำนวณค่าประมาณขนาดประชากรโดยอาศัยข้อมูลที่ได้จากการเลือกตัวอย่างครั้งที่ 1 และ 2 เท่านั้น ด้วยสมการ

$$N_1 = \frac{(n_1+1)(n_2+1)}{(m_2-1)} - 1 \tag{2.3.7}$$

ขั้นตอนที่ 2. : คำนวณค่าประมาณขนาดประชากรโดยอาศัยข้อมูลที่ได้จากการเลือกตัวอย่างครั้งที่ 1 ครั้งที่ 2 และ 3 โดยพิจารณาว่าการเลือกตัวอย่างครั้งที่ 3 เป็นการเลือกตัวอย่างจากประชากรที่มีหน่วยตัวอย่างที่ทำเครื่องหมายแล้วเป็นจำนวน $n_1 + n_2 - m_2$ โดยสมการที่ใช้ในการคำนวณคือ

$$N_2 = \frac{(n_1 + n_2 - m_2 + 1)(n_3+1)}{(m_3+1)} - 1 \tag{2.3.8}$$

ขั้นตอนที่ 3. : คำนวณค่าถ่วงน้ำหนักสำหรับค่าประมาณขนาดประชากร โดยค่าถ่วงน้ำหนักสำหรับ N_1 และ N_2 คือ W_1 และ W_2 ตามลำดับ

$$W_1 = \frac{m_2+1}{m_2+m_3-m_{23}+1} \tag{2.3.9}$$

$$W_2 = \frac{m_3+1}{m_2+m_3-m_{23}+1} \tag{2.3.10}$$

ขั้นตอนที่ 4. : คำนวณค่าประมาณขนาดประชากรสำหรับวิธีนี้โดยสมการต่อไปนี้

$$N_m = W_1N_1 + W_2N_2 \qquad (2.3.11)$$

เมื่อคำนวณค่าประมาณได้แล้ว จะคำนวณค่าร้อยละของความคลาดเคลื่อนสัมพัทธ์สัมบูรณ์จากสมการต่อไปนี้

$$\text{MAPE} = \frac{1}{I}\sum_{i=1}^{I}\left|\frac{(N_i - N_{mi})}{N_i}\right| \times 100 \qquad (2.3.12)$$