บทที่ 3

วิธีการดำเนินการวิจัย

การวิจัยครั้งนี้จะใช้เทคนิคมอนติคาร์โลในการจำลองข้อมูล โดยมีรายละเอียดดังต่อไปนี้

ขั้นตอนในการดำเนินการวิจัย

ขั้นตอนในการดำเนินการวิจัยมีดังนี้

1. สร้างข้อมูลที่มีการแจกแจงแบบเบอร์นูลลี

1.1 กำหนดสัดส่วนประชากร (พารามิเตอร์ p) ตามขอบเขตการวิจัยในข้อ 4

1.2 จำลองข้อมูล ที่มีการแจกแจงแบบเบอร์นูลลีที่มีพารามิเตอร์เท่ากับ p โดยที่จำลองแบบข้อมูลในแต่ละชั้นภูมิ ที่มีการแจกแจงแบบเบอร์นูลลี ทั้งกรณีจำนวนประชากรในแต่ละกลุ่มของแต่ละชั้นภูมิเท่ากันและกรณีจำนวนประชากรในแต่ละกลุ่มของแต่ละชั้นภูมิไม่เท่ากัน ดังนี้

ข้อมูลของชั้นภูมิที่ 1 มีทั้งหมดจำนวน 11 กลุ่ม ได้จากพารามิเตอร์เท่ากับ p_1

ข้อมูลของชั้นภูมิที่ 2 มีทั้งหมดจำนวน 9 กลุ่ม ได้จากพารามิเตอร์เท่ากับ p_2

ข้อมูลของชั้นภูมิที่ 3 มีทั้งหมดจำนวน 12 กลุ่ม ได้จากพารามิเตอร์เท่ากับ p_3

ข้อมูลของชั้นภูมิที่ 4 มีทั้งหมดจำนวน 8 กลุ่ม ได้จากพารามิเตอร์เท่ากับ

$$p_4 = (Np - N_1p_1 - N_2p_2 - N_3p_3) / N_4$$

2. ทำการสุ่มตัวอย่างตามแผนแบบการชักตัวอย่างเชิงซ้อน

2.1 สุ่มกลุ่มซึ่งเป็นหน่วยตัวอย่างขั้นที่ 1 (Primary Sampling Units: PSU) จำนวน 5 กลุ่ม จาก แต่ละชั้นภูมิ

2.2 สุ่มหน่วยตัวอย่างขั้นที่ 2 (Secondary Sampling Units: SSU) ขนาด n_{hi} (กำหนดขนาดตัวอย่างในขอบเขตการวิจัย) ในแต่ละ PSU ที่เป็นตัวอย่างในแต่ละชั้นภูมิ

3. คำนวณหาสัดส่วนของตัวอย่าง $\hat{p}_{(n)}$

สัดส่วนของตัวอย่าง $\hat{p}_{(n)}$ ตามแผนแบบการชักตัวอย่างเกาะกลุ่มสองขั้นแบบแบ่งเป็นชั้น (Stratified Two-Stage Cluster Sampling) คำนวณได้ดังนี้

$\hat{p}_{(n)}$ เป็นค่าประมาณสำหรับพารามิเตอร์ p (สัดส่วนตัวอย่าง)

$$\hat{p}_{(n)} = \frac{\sum_{h=1}^{H} N_h \hat{p}_h}{\sum_{h=1}^{H} N_h}$$

$\hat{p}_h = \frac{1}{N_h}\frac{M_h}{m_h}\sum_{i=1}^{m_h}\frac{N_{hi}}{n_{hi}}\sum_{j=1}^{n_{hi}} x_{hij}$ คือ สัดส่วนตัวอย่างในชั้นภูมิที่ h เมื่อ $h = 1,2,\ldots,H$,

$i = 1,2,\ldots,m_h$, $j = 1,2,\ldots,n_{hi}$

(สูตรนี้ปรับปรุงมาจาก Yamane, 1967 และ Cochran, 1997)

N คือ จำนวนประชากรทั้งหมดในทุกชั้นภูมิ

N_h คือ จำนวนประชากร ทั้งหมดในชั้นภูมิที่ h

H คือ ขนาดชั้นภูมิ

M_h คือ จำนวน PSU ในชั้นภูมิที่ h

m_h คือ จำนวน PSU ที่ถูกสุ่มเป็นตัวอย่างในชั้นภูมิที่ h

N_{hi} คือ จำนวนประชากรของ PSU ที่ถูกสุ่มเลือกเป็นตัวอย่างที่ i ในชั้นภูมิที่ h หรือ คือจำนวน SSU ที่มีอยู่ทั้งหมดใน PSU ที่ถูกสุ่มเลือกเป็นตัวอย่างที่ i ในชั้นภูมิที่ h

n_{hi} คือ จำนวนตัวอย่างที่เลือกจากจำนวนประชากรของ PSU ที่ถูกสุ่มเลือกเป็นตัวอย่างที่ i ในชั้นภูมิที่ h หรือ คือ จำนวน SSU ที่ถูกสุ่มเลือกเป็นตัวอย่างจาก SSU ที่มีอยู่ทั้งหมดใน PSU ถูกสุ่มเลือกเป็นตัวอย่างที่ที่ i ในชั้นภูมิที่ h

x_{hij} คือ ลักษณะภายใต้การศึกษาของตัวอย่างขั้นที่สองหน่วยที่ j ที่สุ่มเลือกจากตัวอย่างขั้นที่หนึ่งหน่วยที่ i ในชั้นภูมิที่ h โดยที่ $x_{hij} = \begin{cases} 1, \text{เกิดลักษณะที่สนใจ} \\ 0, \text{เมื่อเกิดลักษณะที่ไม่สนใจ} \end{cases}$

4. คำนวณช่วงความเชื่อมั่นที่ใช้ในงานวิจัย

ในงานวิจัยครั้งนี้สามารถคำนวณช่วงความเชื่อมั่นจาก 2 วิธีการดังนี้

4.1 วิธีการของวิลสัน (Wilson Method) ที่ใช้ขนาดตัวอย่างแบบเดิม

วิลสัน (Wilson, 1927 อ้างจาก Brown, 2001) ได้เสนอการประมาณค่าแบบช่วงสำหรับสัดส่วนของลักษณะที่สนใจในประชากรโดยมีพื้นฐานมาจากทฤษฎีขีดจำกัดกลาง ซึ่งพิจารณาจากวิธีการหาช่วงความเชื่อมั่นโดยการใช้สถิติทดสอบ อย่างไรก็ตามวิธีการหาช่วง ความเชื่อมั่น มีหลายวิธี เช่น การใช้ปริมาณหมุน การใช้ตัวประมาณแบบจุด การใช้สถิติทดสอบ และ การใช้คุณสมบัติเชิงกำกับของฟังก์ชันควรจะเป็น เป็นต้น

การใช้สถิติทดสอบ การทดสอบสมมติฐานและการประมาณพารามิเตอร์แบบช่วง มีความสัมพันธ์กันอย่างใกล้ชิด จึงอาจหาช่วงความเชื่อมั่น หรือเซตความเชื่อมั่น (Confidence Set) จากเซตหรือบริเวณการยอมรับสมมติฐาน (Acceptance Set) โดยการเปลี่ยนบริเวณยอมรับที่ระดับ α ไปสู่ช่วงหรือเซตความเชื่อมั่นระดับ $1-\alpha$ ได้

ให้ $X_1, X_2, ..., X_n$ เป็นตัวอย่างสุ่มจากประชากรที่มีฟังก์ชันความหนาแน่น $f(x;\theta)$ โดยที่ θ เป็นพามิเตอร์ที่ไม่ทราบค่าซึ่งเป็นจำนวนจริง $(\theta \in \Omega)$ ในการทดสอบสมมติฐาน $H_0 : \theta = \theta_0$ เทียบกับ $H_1 : \theta \neq \theta_0$ เซตย่อย $A(\theta_0)$ ของปริภูมิตัวอย่าง S เป็นบริเวณยอมรับของการทดสอบขนาด α (Acceptance region of level α test) ก็ต่อเมื่อ

$$P[(X_1, ..., X_n) \notin A(\theta_0) \mid H_0 : \theta = \theta_0] \leq \alpha$$

เมื่อกำหนดจุดสังเกต $(x_1, x_2, ..., x_n) \in S$ ให้เซตย่อย $C(x_1, x_2, ..., x_n)$ ของปริภูมิพารามิเตอร์ Ω เป็นเซตความเชื่อมั่น $1-\alpha$ ของ θ ก็ต่อเมื่อ

$$P[C(X_1, X_2, ..., X_n) \in \theta] \geq 1-\alpha \qquad \text{ทุก } \theta \in \Omega$$

(ประชุม สุวัตถี, 2545)

พิจารณาที่ p คือพารามิเตอร์(สัดส่วนของลักษณะที่สนใจในประชากร) p_0 คือปริภูมิตัวอย่าง S เป็นบริเวณยอมรับของการทดสอบขนาด α จะได้ว่าการทดสอบสมมติฐาน $H_0 : p = p_0$ เทียบกับ $H_1 : p \neq p_0$ ที่ระดับนัยสำคัญ α ซึ่งจะยอมรับสมมติฐานหลักเมื่อ

$$\frac{|\hat{p} - p|}{\sqrt{p(1-p)/n}} \quad \leq \quad Z_{1-\alpha/2}$$

ทำการหาช่วงความเชื่อมั่นได้ดังนี้

ยกกำลังสองอสมการข้างต้น จะได้ว่า

$$\left\{\frac{|\hat{p} - p|}{\sqrt{p(1-p)/n}}\right\}^2 \quad \leq \quad Z^2_{1-\alpha/2}$$

$$\frac{(\hat{p} - p)^2}{(\sqrt{p(1-p)/n})^2} \quad \leq \quad Z^2_{1-\alpha/2}$$

$$(\hat{p}-p)^2 \leq Z^2_{1-\alpha/2}\,(\sqrt{p(1-p)/n})^2$$

$$\hat{p}^2-2\hat{p}p+p^2 \leq Z^2_{1-\alpha/2}\,\frac{p(1-p)}{n}$$

$$\hat{p}^2-2\hat{p}p+p^2 \leq Z^2_{1-\alpha/2}\,\frac{p}{n}-Z^2_{1-\alpha/2}\,\frac{p^2}{n}$$

$$\hat{p}^2-2\hat{p}p+p^2-Z^2_{1-\alpha/2}\,\frac{p}{n}+Z^2_{1-\alpha/2}\,\frac{p^2}{n} \leq 0$$

กำหนดให้สมการมีค่าเท่ากับ 0 จะได้

$$\hat{p}^2-p\left(2\hat{p}+\frac{Z^2_{1-\alpha/2}}{n}\right)+p^2\left(1+\frac{Z^2_{1-\alpha/2}}{n}\right) = 0$$

$$p^2\left(1+\frac{Z^2_{1-\alpha/2}}{n}\right)-p\left(2\hat{p}+\frac{Z^2_{1-\alpha/2}}{n}\right)+\hat{p}^2 = 0$$

จะเห็นว่าสมการนี้อยู่ในรูปกำลังสอง (Quadratic equation) คือ

$$ax^2+bx+c = 0$$

นั่นคือ

$$x = p$$

$$a = \left(1+\frac{Z^2_{1-\alpha/2}}{n}\right)$$

$$b = -\left(2\hat{p}+\frac{Z^2_{1-\alpha/2}}{n}\right)$$

$$c = \hat{p}^2$$

และสามารถหารากของสมการได้จากสูตร

$$x = \frac{-b\pm\sqrt{b^2-4ac}}{2a}$$

ดังนั้น

$$p = \frac{\left(2\hat{p}+\frac{Z_{1-\alpha/2}^{2}}{n}\right) \pm \sqrt{\left(2\hat{p}+\frac{Z_{1-\alpha/2}^{2}}{n}\right)^{2} - 4\hat{p}^{2}\left(1+\frac{Z_{1-\alpha/2}^{2}}{n}\right)}}{2\left(1+\frac{Z_{1-\alpha/2}^{2}}{n}\right)}$$

$$p = \frac{2\hat{p}+\frac{Z_{1-\alpha/2}^{2}}{n} \pm \sqrt{4\hat{p}^{2}+4\hat{p}\frac{Z_{1-\alpha/2}^{2}}{n}+\frac{Z_{1-\alpha/2}^{4}}{n^{2}}-4\hat{p}^{2}-4\hat{p}^{2}\frac{Z_{1-\alpha/2}^{2}}{n}}}{2\left(1+\frac{Z_{1-\alpha/2}^{2}}{n}\right)}$$

$$p = \frac{\left(2\hat{p}+\frac{Z_{1-\alpha/2}^{2}}{n}\right) \pm \sqrt{\frac{Z_{1-\alpha/2}^{2}}{n}\left(4\hat{p}(1-\hat{p})+\frac{Z_{1-\alpha/2}^{2}}{n}\right)}}{2\left(1+\frac{Z_{1-\alpha/2}^{2}}{n}\right)}$$

$$p = \frac{\left(\hat{p}+\frac{Z_{1-\alpha/2}^{2}}{2n}\right) \pm \frac{Z_{1-\alpha/2}}{\sqrt{n}}\sqrt{\left(\hat{p}(1-\hat{p})+\frac{Z_{1-\alpha/2}^{2}}{4n}\right)}}{\left(1+\frac{Z_{1-\alpha/2}^{2}}{n}\right)}$$

$$= \left(\frac{\hat{p}n+\frac{Z_{1-\alpha/2}^{2}}{2}}{n+Z_{1-\alpha/2}^{2}}\right) \pm \frac{Z_{1-\alpha/2}\sqrt{n}}{n+Z_{1-\alpha/2}^{2}}\sqrt{\left(\hat{p}(1-\hat{p})+\frac{Z_{1-\alpha/2}^{2}}{4n}\right)}$$

นั่นคือ สามารถคำนวณช่วงความเชื่อมั่นที่ระดับ $100(1-\alpha)\%$ สำหรับ p : $CI_w=(L_w , U_w)$ จากสูตร

$$L_w = \left(\frac{\hat{p}_{(n)}\, n+\frac{Z_{1-\alpha/2}^{2}}{2}}{n+Z_{1-\alpha/2}^{2}}\right) - \frac{Z_{1-\alpha/2}\sqrt{n}}{n+Z_{1-\alpha/2}^{2}}\sqrt{\left(\hat{p}_{(n)}(1-\hat{p}_{(n)})+\frac{Z_{1-\alpha/2}^{2}}{4n}\right)}$$

$$U_w = \left(\frac{\hat{p}_{(n)}\, n+\frac{Z_{1-\alpha/2}^{2}}{2}}{n+Z_{1-\alpha/2}^{2}}\right) + \frac{Z_{1-\alpha/2}\sqrt{n}}{n+Z_{1-\alpha/2}^{2}}\sqrt{\left(\hat{p}_{(n)}(1-\hat{p}_{(n)})+\frac{Z_{1-\alpha/2}^{2}}{4n}\right)}$$

เมื่อ $Z_{1-\alpha/2}$ คือ เปอร์เซ็นไทล์ที่ $(1-\frac{\alpha}{2})100$ ของ N(0,1)

L_w คือ ขีดจำกัดล่างของวิธีการของวิลสันที่ใช้ขนาดตัวอย่างแบบเดิม

U_w คือ ขีดจำกัดบนของวิธีการของวิลสันที่ใช้ขนาดตัวอย่างแบบเดิม

n คือ ขนาดตัวอย่างแบบเดิม

$\hat{p}_{(n)}$ คือ สัดส่วนตัวอย่าง ซึ่งคำนวณจาก

$$\hat{p}_{(n)} = \frac{\sum_{h=1}^{H} N_h \hat{p}_h}{\sum_{h=1}^{H} N_h}$$

$\hat{p}_h = \frac{1}{N_h}\frac{M_h}{m_h}\sum_{i=1}^{m_h}\frac{N_{hi}}{n_{hi}}\sum_{j=1}^{n_{hi}} x_{hij}$ คือ สัดส่วนตัวอย่างในชั้นภูมิที่ h

เมื่อ $h = 1, 2,\ldots, H$, $i= 1, 2, \ldots, m_h$, $j = 1, 2, \ldots, n_{hi}$

(สูตรนี้ปรับปรุงมาจาก Yamane, 1967 และ Cochran, 1997)

N คือ จำนวนประชากรทั้งหมดในทุกชั้นภูมิ

N_h คือ จำนวนประชากร ทั้งหมดในชั้นภูมิที่ h

H คือ ขนาดชั้นภูมิ

M_h คือ จำนวน PSU ที่มีอยู่ทั้งหมดในชั้นภูมิที่ h

m_h คือ จำนวน PSU ที่ถูกสุ่มเลือกเป็นตัวอย่างในชั้นภูมิที่ h

N_{hi} คือ จำนวนประชากรของ PSU ที่ถูกสุ่มเลือกเป็นตัวอย่างที่ i ในชั้นภูมิที่ h หรือคือจำนวน SSU ที่มีอยู่ทั้งหมดใน PSU ที่ถูกสุ่มเลือกเป็นตัวอย่างที่ i ในชั้นภูมิที่ h

n_{hi} คือ จำนวนตัวอย่างที่เลือกจากจำนวนประชากรของ PSU ที่ถูกสุ่มเลือกเป็นตัวอย่างที่ i ในชั้นภูมิที่ h หรือ คือ จำนวน SSU ที่ถูกสุ่มเลือกเป็นตัวอย่างจาก SSU ที่มีอยู่ทั้งหมดใน PSU ถูกสุ่มเลือกเป็นตัวอย่างที่ที่ i ในชั้นภูมิที่ h

x_{hij} คือ ลักษณะภายใต้การศึกษาของตัวอย่างขั้นที่สองหน่วยที่ j ที่สุ่มเลือกจากตัวอย่างขั้นที่หนึ่งหน่วยที่ i ในชั้นภูมิที่ h โดยที่ $x_{hij} = \begin{cases} 1, \text{เกิดลักษณะ ที่สนใจ} \\ 0, \text{เมื่อเกิดลักษณะที่ไม่สนใจ} \end{cases}$

(สูตรนี้ปรับปรุงมาจาก Yamane, 1967 และ Cochran, 1997)

4.2 วิธีการของวิลสันที่ใช้ขนาดตัวอย่างที่มีประสิทธิภาพ

เนื่องจากตัวอย่างที่ใช้ในการทดลองเป็นตัวอย่างที่ได้มาจากเทคนิคการชักตัวอย่างเชิงซ้อนทำให้การค้นพบช่วงความเชื่อมั่นสำหรับสัดส่วนของลักษณะที่สนใจในประชากรที่ใช้กับตัวอย่างเชิงซ้อนนั้นทำได้ยุ่งยากมาก เนื่องจากมีข้อจำกัดบางประการทำให้หน่วยตัวอย่างขั้นที่ 1 (Primary Sampling Units : PSU) ที่เลือกมามีจำนวนน้อย ซึ่งจะมีผลให้องศาแห่งเสรี (Degree of Freedom)มีค่าต่ำ และเนื่องจากตัวประมาณค่าที่ดีจะขึ้นอยู่กับความคลาดเคลื่อนในการสุ่มตัวอย่างของแต่ละ PSU เมื่อ PSUมีจำนวนน้อยทำให้ความคลาดเคลื่อนของตัวประมาณผิดปกติ (Oranji , 2006) ซึ่งทั้งสองสาเหตุสามารถแก้ไขได้ โดยคอร์นและกรูบาร์ด (Korn and Graubard ,1998) อ้างจาก Gray,Haslett and Kuzmicich, 2004)และ ออเรนจิ (Oranji, 2006) แนะนำให้หาขนาดตัวอย่างใหม่โดยประมาณขนาดตัวอย่างที่มีประสิทธิภาพ (Effective Sample Size) แทนด้วย $\tilde{n}$ โดยหาอัตราส่วนระหว่างขนาดตัวอย่างแบบเดิมกับอิทธิพลของแผนแบบ (Design Effect : DEFF) โดยที่สามารถคำนวณได้ตามขั้นตอนดังนี้

4.2.1 คำนวณค่าอิทธิพลของแผนแบบ (Design Effect): DEFF

$$\text{DEFF} = \frac{V_{\text{complex}}}{V_{\text{SRS}}}$$

โดยที่ V_{SRS} คือความแปรปรวนของสัดส่วนตัวอย่างที่ได้มาจากแผนแบบการชักตัวอย่างแบบสุ่มเชิงเดียว (Simple Random Sampling)

V_{complex} คือ ความแปรปรวนของสัดส่วนตัวอย่างเชิงซ้อนซึ่งจะคำนวณจากวิธีการหาตัวประมาณแบบแจ็คไนฟ์ หรืออาจแทนด้วย V_{JK}

4.2.1.1 คำนวณหาตัวประมาณของ V_{SRS} $(\hat{V}_{\text{SRS}})$ ดังนี้

$$\hat{V}_{\text{SRS}} = \frac{\hat{p}_{\text{SRS}}(1-\hat{p}_{\text{SRS}})}{\sum_{h=1}^{H}\sum_{i=1}^{m_h} n_{hi}}$$

เมื่อ $\hat{p}_{\text{SRS}}$ คือ สัดส่วนของตัวอย่างตามแผนแบบการชักตัวอย่างแบบสุ่มเชิงเดียวซึ่งคำนวณจาก $\hat{p}_{\text{SRS}} = \dfrac{\sum_{h=1}^{H}\sum_{i=1}^{m_h}\sum_{j=1}^{n_{hi}} x_{hij}}{\sum_{h=1}^{H}\sum_{i=1}^{m_h} n_{hi}}$

โดยที่ $\sum_{h=1}^{H}\sum_{i=1}^{m_h}\sum_{j=1}^{n_{hi}} x_{hij}$ คือ ผลรวมของลักษณะภายใต้การศึกษาของตัวอย่างขั้นที่สองหน่วยที่ j ที่สุ่มเลือกจากตัวอย่างขั้นที่หนึ่งหน่วยที่ i ในชั้นภูมิที่ h

$\sum_{h=1}^{H}\sum_{i=1}^{m_h} n_{hi}$ คือ จำนวนตัวอย่างทั้งหมดที่ทำการศึกษา

4.2.1.2 คำนวณหาตัวประมาณของ $V_{complex}$ ($\hat{V}_{complex}(\hat{p}_{(n)})$) หรือ V_{JK} ($\hat{V}_{JK}(\hat{p}_{(n)})$) จากวิธีการหาตัวประมาณแบบแจ็คไนฟ์ดังนี้

ถ้าให้ p เป็นพารามิเตอร์ (สัดส่วนของลักษณะที่สนใจในประชากร)

$\hat{p}_{(n)}$ เป็นค่าประมาณสำหรับพารามิเตอร์ p (สัดส่วนตัวอย่าง)

จาก
$$\hat{p}_{(n)} = \frac{\sum_{h=1}^{H} N_h \hat{p}_h}{\sum_{h=1}^{H} N_h}$$

จะได้ว่า
$$V(\hat{p}_{(n)}) = \frac{\sum_{h=1}^{H} N_h^{\,2} V(\hat{p}_h)}{\left(\sum_{h=1}^{H} N_h\right)^2}$$

ดังนั้น ตัวประมาณค่าความแปรปรวนโดยวิธีแจ็คไนฟ์ของ $\hat{p}_{(n)}$ คือ

$$\hat{V}_{JK}(\hat{p}_{(n)}) = \frac{\sum_{h=1}^{H} N_h^{\,2} \hat{V}_{JK}(\hat{p}_h)}{\left(\sum_{h=1}^{H} N_h\right)^2}$$

เมื่อ $\hat{V}_{JK}(\hat{p}_h)$ คือ ตัวประมาณค่าความแปรปรวนโดยวิธีแจ็คไนฟ์ของ $\hat{p}_h$ ที่อยู่ในชั้นภูมิที่ h ซึ่งในแต่ละชั้นภูมิมีการชักตัวอย่างแบบเกาะกลุ่มสองขั้น (Two-Stage Cluster Sampling) ทำให้สามารถหา $\hat{V}_{JK}(\hat{p}_h)$ จากกรณีที่ m>1 (ดูจากหัวข้อที่ 1.3 ในบทที่ 2)

4.2.2 คำนวณขนาดตัวอย่างที่มีประสิทธิภาพ ($\tilde{n}$) ได้จากสูตร

$$\tilde{n} = \frac{n}{DEFF}$$

เมื่อ $\sum_{h=1}^{H}\sum_{i=1}^{m_h} n_{hi} = n$

ดังนั้นสามารถคำนวณหาช่วงความเชื่อมั่นของสัดส่วนของลักษณะที่สนใจในประชากรสำหรับการชักตัวอย่างเชิงซ้อนโดยดัดแปลงวิธีการของวิลสัน สามารถทำได้ดังนี้

ช่วงความเชื่อมั่นที่ระดับ $100(1-\alpha)\%$ สำหรับ p : CI_{wc} = (L_{wc} , U_{wc}) คือ

$$L_{wc} = \left(\frac{\hat{p}_{(\tilde{n})}\tilde{n} + \frac{Z^2_{1-\alpha/2}}{2}}{\tilde{n} + Z^2_{1-\alpha/2}} \right) - \frac{Z_{1-\alpha/2}\sqrt{\tilde{n}}}{\tilde{n} + Z^2_{1-\alpha/2}} \sqrt{\left(\hat{p}_{(\tilde{n})}(1-\hat{p}_{(\tilde{n})}) + \frac{Z^2_{1-\alpha/2}}{4\tilde{n}} \right)}$$

$$U_{wc} = \left(\frac{\hat{p}_{(\tilde{n})}\tilde{n} + \frac{Z^2_{1-\alpha/2}}{2}}{\tilde{n} + Z^2_{1-\alpha/2}} \right) + \frac{Z_{1-\alpha/2}\sqrt{\tilde{n}}}{\tilde{n} + Z^2_{1-\alpha/2}} \sqrt{\left(\hat{p}_{(\tilde{n})}(1-\hat{p}_{(\tilde{n})}) + \frac{Z^2_{1-\alpha/2}}{4\tilde{n}} \right)}$$

เมื่อ

$Z_{1-\alpha/2}$ คือ เปอร์เซ็นไทล์ที่ $(1-\frac{\alpha}{2})100$ ของ N(0,1)

L_w คือ ขีดจำกัดล่างของวิธีการของวิลสันที่ใช้ขนาดตัวอย่างที่มีประสิทธิภาพ

U_w คือ ขีดจำกัดบนของวิธีการของวิลสันที่ใช้ขนาดตัวอย่างที่มีประสิทธิภาพ

$\tilde{n}$ คือ ขนาดตัวอย่างที่มีประสิทธิภาพ

$\hat{p}_{(\tilde{n})}$ คือ สัดส่วนตัวอย่างที่ใช้ขนาดตัวอย่างที่มีประสิทธิภาพ ซึ่งคำนวณจาก

$$\hat{p}_{(\tilde{n})} = \frac{\sum_{h=1}^{H} N_h \hat{p}_h}{\sum_{h=1}^{H} N_h}$$

$\hat{p}_h = \frac{1}{N_h}\frac{M_h}{m_h}\sum_{i=1}^{m_h}\frac{N_{hi}}{\tilde{n}_{hi}}\sum_{j=1}^{\tilde{n}_{hi}} x_{hij}$ คือ สัดส่วนตัวอย่างในชั้นภูมิที่ h

เมื่อ h = 1,2,…,H , i = 1,2,…,m_h , j = 1,2,…, $\tilde{n}_{hi}$

(สูตรนี้ปรับปรุงมาจาก Yamane, 1967 และ Cochran, 1997)

N คือ จำนวนประชากรทั้งหมดในทุกชั้นภูมิ

N_h คือ จำนวนประชากร ทั้งหมดในชั้นภูมิที่ h

H คือ ขนาดชั้นภูมิ

M_h คือ จำนวน PSU ที่มีอยู่ทั้งหมดในชั้นภูมิที่ h

m_h คือ จำนวน PSU ที่ถูกสุ่มเลือกเป็นตัวอย่างในชั้นภูมิที่ h

N_{hi} คือ จำนวนประชากรของ PSU ที่ถูกสุ่มเลือกเป็นตัวอย่างที่ i ในชั้นภูมิที่ h หรือคือจำนวน SSU ที่มีอยู่ทั้งหมดใน PSU ที่ถูกสุ่มเลือกเป็นตัวอย่างที่ i ในชั้นภูมิที่ h

$\tilde{n}_{hi}$ คือ จำนวนตัวอย่างที่รวมตัวอย่างที่เลือกใหม่จากจำนวนประชากรของ PSU ที่ถูกสุ่มเลือกเป็นตัวอย่างที่ i ในชั้นภูมิที่ h หรือ คือ จำนวน SSU ที่ถูกสุ่มเลือกเป็นตัวอย่างจาก SSU ที่มีอยู่ทั้งหมดใน PSU ถูกสุ่มเลือกเป็นตัวอย่างที่ที่ i ในชั้นภูมิที่ h

x_{hij} คือ ลักษณะภายใต้การศึกษาของตัวอย่างขั้นที่สองหน่วยที่ j ที่สุ่มเลือกจากตัวอย่างขั้นที่หนึ่งหน่วยที่ i ในชั้นภูมิที่ h โดยที่ $x_{hij} = \begin{cases} 1, \text{เกิดลักษณะ ที่สนใจ} \\ 0, \text{เมื่อเกิดลักษณะที่ไม่สนใจ} \end{cases}$

ทำซ้ำข้อ 1 ถึง 4 จำนวน 1,000 ครั้ง ในแต่ละสถานการณ์

5. คำนวณเกณฑ์ที่ใช้ในการเปรียบเทียบ

5.1 ค่าประมาณสัมประสิทธิ์ความเชื่อมั่น (Confidence Coefficient Estimator) ค่าประมาณสัมประสิทธิ์ความเชื่อมั่น $(1-\hat{\alpha})$ คือค่าความน่าจะเป็นที่ช่วงความเชื่อมั่นที่คำนวณได้จะครอบคลุมพารามิเตอร์ p โดยคำนวณได้จาก

$$1-\hat{\alpha} = \frac{\left[\begin{array}{l}\text{จำนวนครั้ง ทั้งหมดที่ ช่วงความเชื่อมั่น} \\ \text{ครอบคลุมพา รามิเตอร์ p}\end{array}\right]}{R}$$

เมื่อ R คือ จำนวนรอบของการทดลอง

5.2 คำนวณความกว้างเฉลี่ยของช่วงความเชื่อมั่น ในกรณีที่ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นอยู่ในขอบเขตที่เชื่อมั่นได้ของค่าสัมประสิทธิ์ความเชื่อมั่น เมื่อสัดส่วนของตัวอย่างที่ใช้ขนาดตัวอย่างแบบเดิม $(\hat{p}_n)$ ตามแผนแบบการชักตัวอย่างเชิงซ้อน และ สัดส่วนของตัวอย่างที่ใช้ขนาดตัวอย่างที่มีประสิทธิภาพ $(\hat{p}_{\tilde{n}})$ ตามแผนแบบการชักตัวอย่างเชิงซ้อน

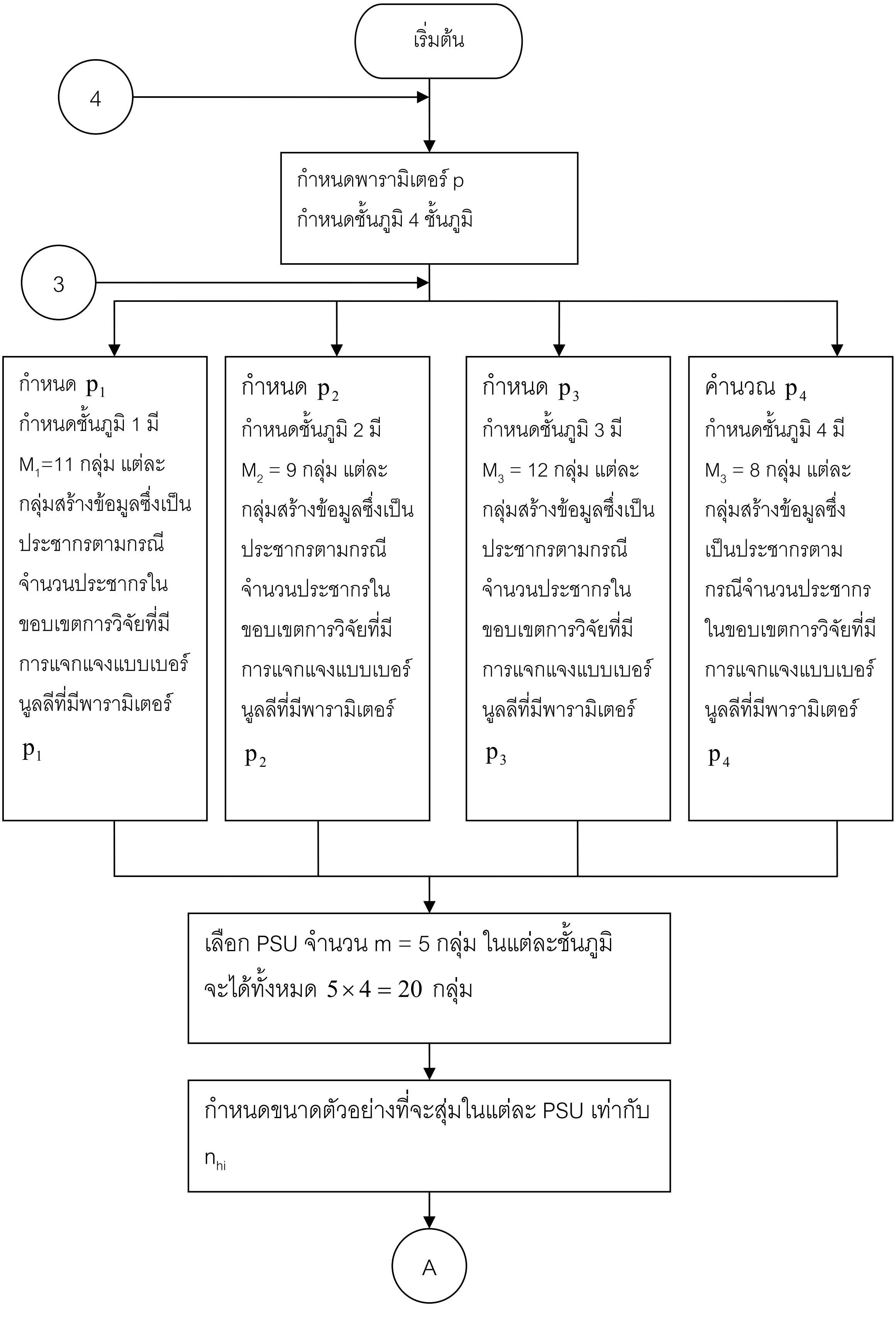
$$AL = \frac{\sum_{j=1}^{R}(U_j - L_j)}{R} \quad ; \quad j=1, 2, \ldots, R$$

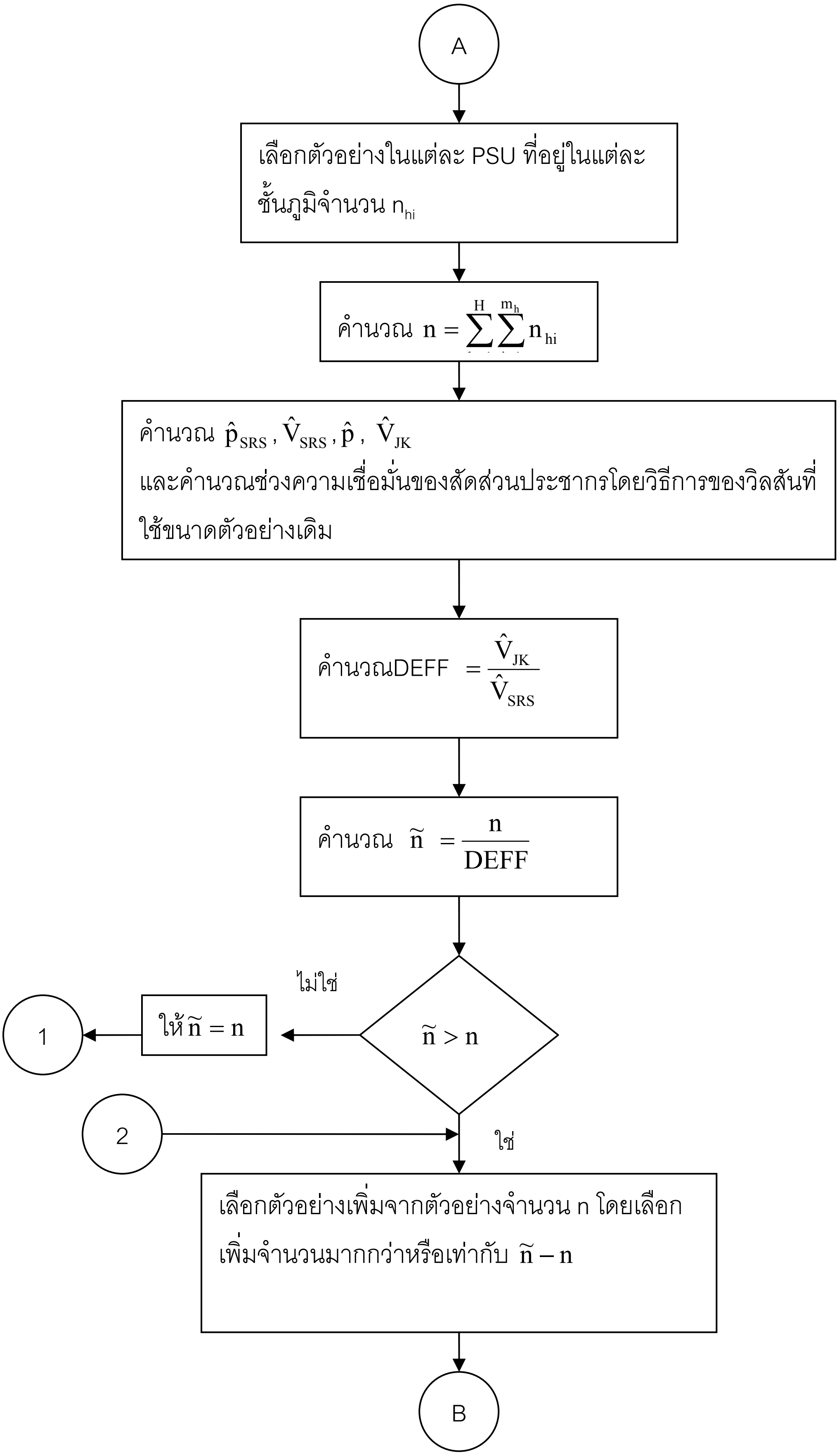
เมื่อ R คือ จำนวนรอบของการทดลอง

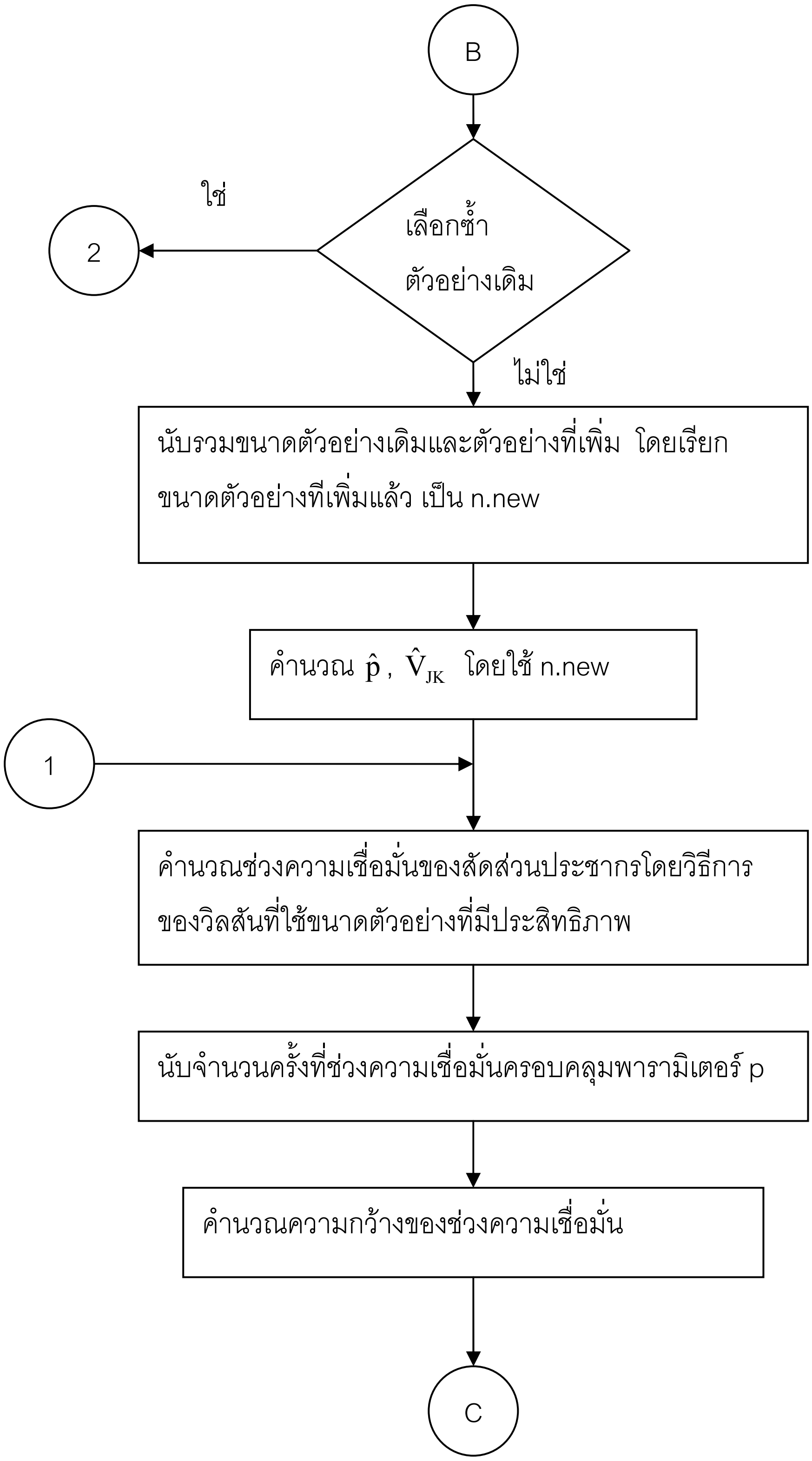
L_j คือ ขีดจำกัดล่างของความเชื่อมั่นของรอบที่ j

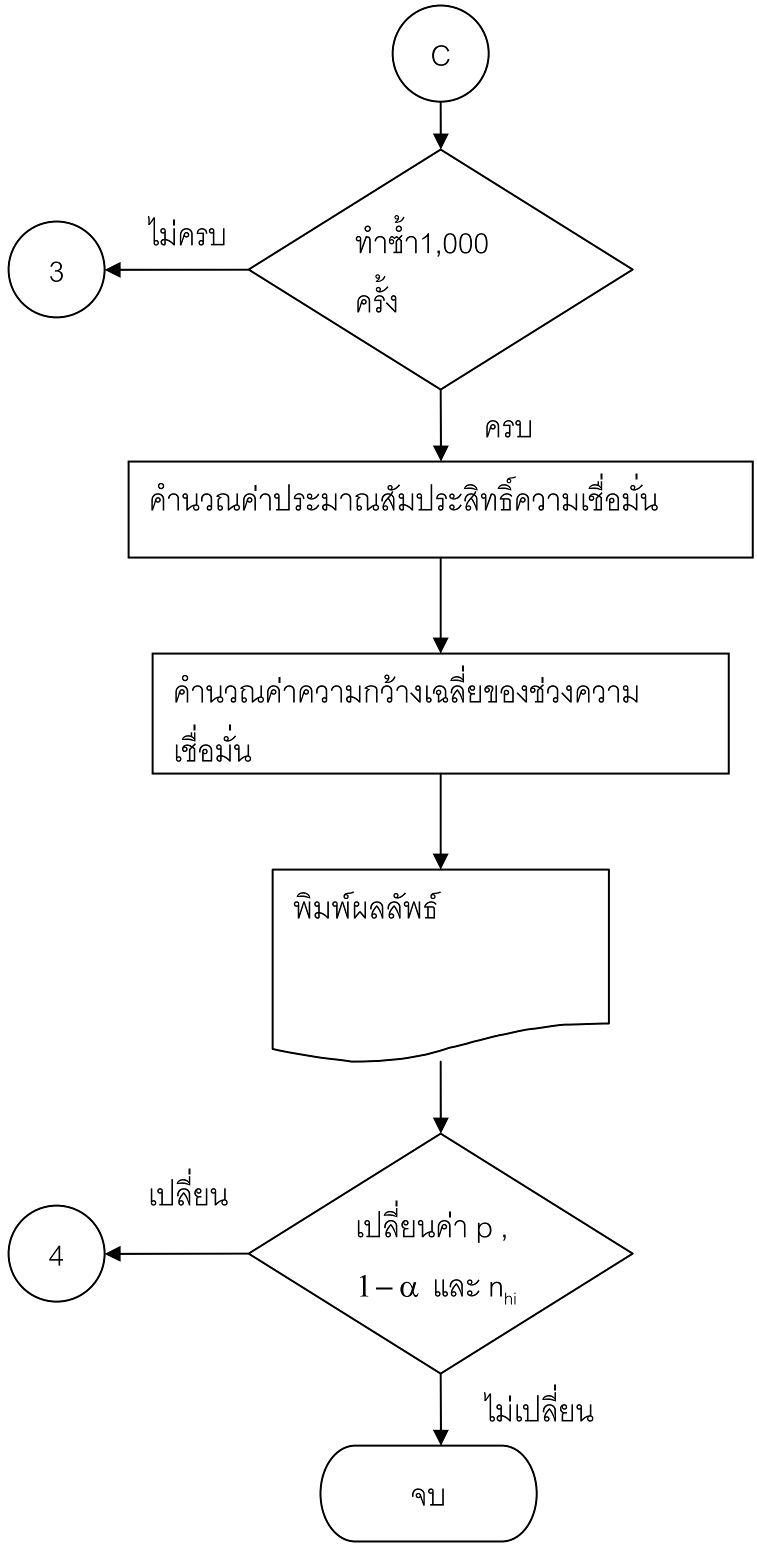
U_j คือ ขีดจำกัดบนของความเชื่อมั่นของรอบที่ j

แผนภาพแสดงขั้นตอนการดำเนินการวิจัย




A
เลือกตัวอย่างในแต่ละ PSU ที่อยู่ในแต่ละชั้นภูมิจำนวน n_{hi}
คำนวณ $n = \sum^{H} \sum^{m_h} n_{hi}$
คำนวณ $\hat{p}_{SRS}$, $\hat{V}_{SRS}$, $\hat{p}$, $\hat{V}_{JK}$
และคำนวณช่วงความเชื่อมั่นของสัดส่วนประชากรโดยวิธีการของวิลสันที่ใช้ขนาดตัวอย่างเดิม
คำนวณDEFF $= \frac{\hat{V}_{JK}}{\hat{V}_{SRS}}$
คำนวณ $\tilde{n} = \frac{n}{DEFF}$
$\tilde{n} > n$
ไม่ใช่
ให้ $\tilde{n} = n$
1
2
ใช่
เลือกตัวอย่างเพิ่มจากตัวอย่างจำนวน n โดยเลือกเพิ่มจำนวนมากกว่าหรือเท่ากับ $\tilde{n} - n$
B


B
ใช่
เลือกซ้ำ
ตัวอย่างเดิม
2
ไม่ใช่
นับรวมขนาดตัวอย่างเดิมและตัวอย่างที่เพิ่ม โดยเรียก
ขนาดตัวอย่างทีเพิ่มแล้ว เป็น n.new
คำนวณ $\hat{p}$, $\hat{V}_{JK}$ โดยใช้ n.new
1
คำนวณช่วงความเชื่อมั่นของสัดส่วนประชากรโดยวิธีการ
ของวิลสันที่ใช้ขนาดตัวอย่างที่มีประสิทธิภาพ
นับจำนวนครั้งที่ช่วงความเชื่อมั่นครอบคลุมพารามิเตอร์ p
คำนวณความกว้างของช่วงความเชื่อมั่น
C



แผนภาพที่ 3.1 แสดงขั้นตอนการดำเนินการวิจัย