บทที่ 2

ตัวสถิติที่ใช้ในการวิจัย

ในการวิจัยครั้งนี้ มีวิธีประมาณขนาดประชากรที่แตกต่างกัน 3 วิธี โดยในแต่ละวิธีจะมีการคำนวณที่แตกต่างกันดังนี้

1. ตัวประมาณ Petersen เป็นตัวประมาณที่ได้จากการเลือกตัวอย่างสองครั้ง และพิจารณาถึงจำนวนตัวอย่างที่ซ้ำกัน โดยครั้งแรกจะเลือกตัวอย่าง n_1 หน่วยจากประชากรขนาด N หน่วย และทำเครื่องหมายหน่วยตัวอย่างที่ถูกเลือก จากนั้นปล่อยหน่วยตัวอย่างกลับไปปะปนกับประชากรที่เหลือ ทำการเลือกตัวอย่างครั้งที่สอง n_2 หน่วย โดยในครั้งนี้ให้ทำการบันทึกจำนวนหน่วยตัวอย่างที่ถูกทำเครื่องหมายจากการเลือกครั้งแรกไว้ด้วย ให้ m_2 คือ จำนวนหน่วยตัวอย่างที่ถูกทำเครื่องหมายจากการเลือกครั้งแรกและถูกเลือกในครั้งที่สอง จะได้ว่า m_2 มีการแจกแจงแบบไฮเปอร์จีออเมตริก (Hypergeometric Distribution) ซึ่งมีฟังก์ชั่นความน่าจะเป็นดังนี้

$$f(m_2|n_1,n_2) = \frac{\binom{n_1}{m_2}\binom{N-n_1}{n_2-m_2}}{\binom{N}{m_2}} \tag{1}$$

โดยที่ $\max(0, n_1+n_2-N) \leq m_2 \leq \min(n_1, n_2)$

หากทำการประมาณขนาดประชากรด้วยตัวประมาณซึ่งมีค่าเท่ากับ $\frac{n_1 n_2}{m_2}$ จะได้ว่ามีความเอนเอียงค่อนข้างสูง ดังนั้น Robson และ Regier (1964) จึงได้ทำการเปลี่ยนแปลงตัวประมาณขึ้นใหม่เป็น

$$\hat{N}_p = \frac{(n_1+1)(n_2+1)}{(m_2+1)} - 1 \tag{2}$$

ซึ่งตัวประมาณนี้ไม่เอนเอียงเมื่อ $n_1 + n_2 \geq N$ และเกือบจะไม่เอนเอียงในกรณีอื่น ๆ

2. ตัวประมาณที่อาศัย Gibbs Sampler จะมีวิธีการเลือกตัวอย่างคล้ายกับตัวประมาณแรก แต่จะทำการเลือกตัวอย่างมากกว่า 2 ครั้ง ทำให้การหารูปแบบของการแจกแจงหลังการทดลอง (Posterior Distribution) ทำได้ยาก ดังนั้นจึงนำวิธี Gibbs Sampler มาช่วยในการหารูปแบบการแจกแจง และทำการประมาณขนาดประชากรจากการแจกแจงที่ได้นั้น

ถ้ากำหนดให้

N = ขนาดประชากร

n_i = จำนวนหน่วยตัวอย่างที่ถูกเลือกในครั้งที่ i ; $i = 1, \ldots, I$

p_i = ความน่าจะเป็นที่หน่วยตัวอย่างจะถูกเลือกในครั้งที่ i ; $i = 1, \ldots, I$

m_i = จำนวนหน่วยตัวอย่างที่มีเครื่องหมายที่ถูกเลือกในครั้งที่ i ; $m_1 = 0$, $i = 1, \ldots, I$

r = จำนวนหน่วยตัวอย่างทั้งหมดที่ถูกทำเครื่องหมาย

$$= \sum_{i=1}^{I} n_i - \sum_{i=1}^{I} m_i$$

จากฟังก์ชันการแจกก่อนการทดลอง (Prior Distributions) $\pi(N, p)$ อาศัย Gibbs Sampler ทำการประมาณ marginal posterior ของ N โดยการจำลองตัวอย่างจากการแจกแจงแบบมีเงื่อนไขของ N และ P ซึ่งจะได้ลำดับดังนี้

$$N^{(0)}, p^{(0)}, N^{(1)}, p^{(1)}, \ldots,$$

โดยที่ $N^{(k)} \sim \pi(N \mid p^{(k-1)}, D)$ และ $p^{(k)} \sim \pi(p \mid N^{(k)}, D)$

จากการกำหนดให้ $\pi(N) = 1/N$ และ $\pi(p_i) \sim \text{Beta}(a,b)$ สำหรับ $i = 1, 2, 3$ จะได้ว่า การแจกแจงหลังการทดลองแบบมีเงื่อนไขของ N คือ การแจกแจงแบบ Negative Binomial ที่มีพารามิเตอร์เป็น $(r-1, 1-\prod_{i=1}^{I}(1-p_i))$ และการแจกแจงหลังการทดลองแบบมีเงื่อนไขของ p_i คือการแจกแจงแบบ Beta ที่มีพารามิเตอร์เป็น $(n_i+a, N-n_i+b)$ สำหรับ $i = 1, 2, 3$ เมื่อทำการจำลองลำดับของ N และ p ไปเป็นจำนวน m รอบ จะนำค่า $N^{(m)}$ ไปเป็นค่าเริ่มต้นสำหรับทำการจำลองในครั้งต่อไป ถ้าทำการจำลองทั้งหมด k ชุด จะได้ว่าตัวประมาณ N คือ

$$\hat{N}_g = \sum_{k=1}^{K} \frac{N^{(k)}}{K} \tag{3}$$

3. ตัวประมาณที่พัฒนาจาก Petersen เป็นตัวประมาณที่ผู้ทำการวิจัยพัฒนาขึ้น ซึ่งจะมีวิธีในการประมาณคล้ายคลึงกับวิธีที่ใช้สำหรับตัวประมาณ Petersen แต่แตกต่างกันตรงที่ตัวประมาณนี้จะต้องมีการทำการเลือกตัวอย่าง 3 ครั้ง และจะต้องทราบว่าเครื่องหมายที่พบในหน่วยตัวอย่างเป็นเครื่องหมายที่ได้จากการเลือกตัวอย่างในครั้งใด (ครั้งที่ 1 หรือ 2) จากนั้นทำการประมาณขนาดประชากรโดยอาศัยข้อมูลที่ได้จากการเลือกตัวอย่างครั้งที่ 1 และ 2 ซึ่งจะให้ค่าประมาณเท่ากับวิธี Petersen แล้วทำการประมาณขนาดประชากรที่ได้หลังจากการเลือกตัวอย่างครั้งที่ 3 โดยใช้สูตรเดียวกัน แต่พิจารณาว่าขนาดตัวอย่างที่มีเครื่องหมายเท่ากับ $n_1+n_2-m_2$

ถ้ากำหนดให้

N = ขนาดประชากร

n_i = จำนวนหน่วยตัวอย่างที่ถูกเลือกในครั้งที่ i เมื่อ $i = 1, 2, 3$

m_j = จำนวนหน่วยตัวอย่างที่มีเครื่องหมายและถูกเลือกในครั้งที่ j เมื่อ $j = 2, 3$

m_{23} = จำนวนหน่วยตัวอย่างที่มีเครื่องหมายและถูกเลือกในครั้งที่ 2 และ 3

$\hat{N}_j$ = ค่าประมาณขนาดประชากรที่ได้จากการเลือกตัวอย่างภายหลังครั้งที่ j เมื่อ $j = 2, 3$

จะได้ว่า หลังจากการเลือกตัวอย่าง 2 ครั้ง ตัวประมาณที่ได้คือ

$$\hat{N}_2 = \frac{(n_1+1)(n_2+1)}{(m_2+1)} - 1 \qquad (4)$$

ส่วนตัวประมาณภายหลังจากการเลือกตัวอย่างครั้งที่ 3 คือ

$$\hat{N}_3 = \frac{(n_1+n_2-m_2+1)(n_3+1)}{(m_3+1)} - 1 \qquad (5)$$

และตัวประมาณที่พัฒนาจาก Petersen คือ

$$\hat{N}_m = \left(\frac{m_2+1}{m_2+m_3-m_{23}+1}\right)\hat{N}_2 + \left(\frac{m_3+1}{m_2+m_3-m_{23}+1}\right)\hat{N}_3 \qquad (6)$$