บทที่ 1

บทนำ

ที่มาและความสำคัญของปัญหา

การประมาณค่า เป็นวิธีการหนึ่งที่สามารถทำให้มีข้อมูลเกี่ยวกับเรื่องที่สนใจได้มากขึ้น ทั้งนี้การประมาณค่าจะต้องอาศัยข้อมูลเพื่อใช้ในการประมาณ โดยข้อมูลนี้อาจได้มาจากการวางแผนการทดลองหรือการสำรวจ ซึ่งการวางแผนการทดลองหรือการสำรวจนี้มีได้หลายวิธี เช่น การวางแผนการทดลองแบบสุ่มบล็อกอย่างง่าย (Randomize Blook Design) , การวางแผนการทดลองแบบลาตินสแควร์ (Latin Square Design) หรือการเลือกตัวอย่างแบบสุ่มอย่างง่าย (Simple Random Sampling) , การเลือกตัวอย่างแบบโควต้า (Quota Sampling) เป็นต้น ในแต่ละวิธีก็จะมีการประมาณค่าที่แตกต่างกันไป ทั้งนี้วิธีการประมาณแบบต่าง ๆ ก็จะเหมาะสมสำหรับลักษณะข้อมูลและตัวแปรที่สนใจประมาณแตกต่างกัน ดังนั้นจึงควรมีการพิจารณาเพื่อเลือกใช้ให้เหมาะสม

ในการวิจัยนี้ ผู้ทำการวิจัยมีความสนใจที่จะประมาณค่าขนาดประชากร เพราะขนาดประชากรมีความสำคัญสำหรับหลาย ๆ เรื่อง เช่น ขนาดประชากรของสัตว์ป่า ขนาดประชากรของผู้ติดยาเสพติด หรือขนาดประชากรของผู้ที่เป็นโรคเอดส์ ทั้งนี้เพื่อที่จะสามารถวางแผนในการอนุรักษ์หรือป้องกันและควบคุมขนาดประชากรได้ แต่จะสังเกตได้ว่าจากประชากรที่กล่าวข้างต้นนั้น เราสามารถรู้ขนาดประชากรที่แท้จริงได้ยาก จึงต้องทำการประมาณขนาดประชากรขึ้น วิธีหนึ่งที่ทำได้และน่าสนใจ คือ การประมาณขนาดประชากรโดยอาศัยข้อมูลที่ได้จากการเลือกตัวอย่างแบบแคบเจอร์-รีแคบเจอร์ (Capture-recapture)

ในปี ค.ศ. 1986 Laplace ได้นำเอาวิธีเลือกตัวอย่างแบบแคบเจอร์-รีแคบเจอร์ มาใช้ในการประมาณขนาดประชากรฝรั่งเศส ต่อมาวิธีการนี้ได้รับความสนใจและนำไปใช้ประมาณขนาดประชากรสัตว์ป่า วิธีนี้มีขั้นตอนไม่ซับซ้อนนัก คือ ทำการเลือกตัวอย่างครั้งที่ 1 และทำเครื่องหมายบนหน่วยตัวอย่างที่เลือกได้ จากนั้นปล่อยหน่วยตัวอย่างที่ทำเครื่องหมายแล้วส่งกลับคืนประชากร ทิ้งช่วงระยะเวลานานพอที่หน่วยตัวอย่างที่ปล่อยไปจะปะปนกับประชากรอย่างทั่วถึง ทำการเลือกตัวอย่างใหม่อีกและทำการบันทึกจำนวนหน่วยตัวอย่างที่มีเครื่องหมาย และทำเครื่อง

หมายเพิ่มสำหรับหน่วยที่ไม่มีเครื่องหมาย แล้วส่งกลับคืนประชากร ทำเช่นนี้ซ้ำ ๆ กันตามที่ต้องการ ก็จะได้ข้อมูลที่ใช้ในการประมาณขนาดประชากรต่อไป สำหรับการเลือกตัวอย่างในประชากรสัตว์ป่า อาจจะทำได้โดยการศึกษาพฤติกรรมของสัตว์ว่ามีเส้นทางการเดินอย่างไรหรือมีลักษณะการหากินอย่างไร แล้วจึงทำการวางกับดักจับสัตว์นั้น ๆ เมื่อจับได้แล้วจึงทำเครื่องหมายให้กับสัตว์ ส่วนการเลือกตัวอย่างของประชากรผู้ติดยาเสพติด อาจจะทำได้โดยได้เลือกตัวอย่างจากโรงพยาบาลที่มีผู้ติดยาเสพติดไปรับการรักษา ทำการบันทึกรายชื่อของผู้ติดยาที่พบจากนั้นไปเลือกตัวอย่างจากสถานีตำรวจที่จับผู้ติดยาเสพติดว่ามีรายชื่อตรงกันหรือไม่ ทำการบันทึกจำนวนรายชื่อผู้ติดยาที่พบซ้ำกัน เป็นต้น แต่จากการเลือกตัวอย่างแบบแคบเจอร์-รีแคบเจอร์นี้ เราสามารถประมาณขนาดประชากรได้จากตัวประมาณหลายตัว ซึ่งวิธีการคำนวณตัวประมาณแต่ละตัวนั้นแตกต่างกันไป มีทั้งยากและง่าย มีทั้งตัวประมาณที่ให้ความถูกต้องแม่นยำสูงและต่ำ การจะเลือกใช้ตัวประมาณใดจึงควรพิจารณาถึงสิ่งเหล่านี้ด้วย ตัวประ-มาณที่น่าสนใจตัวหนึ่ง คือ ตัวประมาณ Petersen ซึ่งเป็นตัวประมาณที่คำนวณได้ง่าย ตัวประ-มาณอีกตัวหนึ่ง คือ ตัวประมาณที่อาศัย Gibbs Sampler ซึ่งมีวิธีการคำนวณค่อนข้างยุ่งยาก แต่อาจจะให้ค่าประมาณที่ดีกว่า นอกจากนี้ยังมีตัวประมาณที่ผู้วิจัยพัฒนาจากตัวประมาณ Petersen จึงสนใจที่จะทำการเปรียบเทียบตัวประมาณทั้งสามว่า ตัวใดจะสามารถประมาณขนาดประชากรได้ดีกว่ากันและดีกว่าในกรณีใด

วัตถุประสงค์ของการวิจัย

เพื่อศึกษาและเปรียบเทียบความถูกต้องของตัวประมาณทั้ง 3 ที่นำมาใช้ประมาณขนาดประชากร โดยอาศัยข้อมูลที่ได้จากการเลือกตัวอย่างแบบแคบเจอร์-รีแคบเจอร์

สมมติฐานของการวิจัย

1. ตัวประมาณที่พัฒนาจากตัวประมาณ Petersen จะประมาณได้ดีกว่าตัวประมาณ Petersen ในกรณีที่ประชากรมีขนาดปานกลางและในกรณีที่ร้อยละของขนาดตัวอย่างมีค่าต่ำ ๆ

2. ตัวประมาณที่พัฒนาจาก Petersen จะประมาณได้ดีกว่าตัวประมาณที่อาศัย Gibbs Sampler ในกรณีที่ประชากรมีขนาดปานกลาง

ข้อตกลงเบื้องต้น

สำหรับการวิจัยครั้งนี้ จะมีข้อตกลงเบื้องต้น ดังนี้

1. ประชากรเป็นประชากรปิด (Closed Population) คือ มีขนาดประชากรคงที่ (N constant)

2. ในการเลือกตัวอย่างแต่ละครั้ง ทุกหน่วยจะมีโอกาสในการถูกเลือกเท่า ๆ กัน นั่นคือ $p_{ij} = p_i$ เมื่อ $j = 1, \ldots, N$ และ $i = 1, \ldots, I$

3. การทำเครื่องหมายให้หน่วยที่ถูกเลือกเป็นตัวอย่างในแต่ละครั้งไม่มีผลต่อการเลือกตัวอย่างในครั้งต่อไป

4. การเลือกตัวอย่างในแต่ละครั้งเป็นไปอย่างสุ่ม และเป็นอิสระต่อกัน

5. หน่วยตัวอย่างไม่ได้ทำเครื่องหมายสูญหาย ในระหว่างการเลือกตัวอย่างสองครั้ง

ขอบเขตการวิจัย

ในการวิจัยครั้งนี้ ได้กำหนดขอบเขตของการวิจัยไว้ดังนี้

1. กำหนดขนาดประชากรที่ต้องการทำการประมาณเป็น 2 ขนาด คือ

   1.1 ประชากรขนาดเล็ก ได้แก่ N = 1000 , 3000 , 5000 , 7000

   1.2 ประชากรขนาดปานกลาง ได้แก่ N = 10000 , 30000 , 50000 , 70000

2. กำหนดขนาดตัวอย่างที่ทำการเลือกในครั้งแรก เป็นร้อยละของขนาดประชากร โดยมีค่าร้อยละดังนี้คือ 1% , 2% , 3% , 4% , 5% , 6% , 7% , 8% , 9% , 10%

3. กำหนดจำนวนหน่วยตัวอย่างที่ถูกเลือกซ้ำในการเลือกครั้งที่ 2 และ 3 ดังนี้

   3.1 ให้จำนวนหน่วยตัวอย่างที่ถูกเลือกซ้ำเป็นไปตามทฤษฎีความน่าจะเป็น

   3.2 ให้จำนวนหน่วยตัวอย่างที่ถูกเลือกซ้ำในการเลือกครั้งที่ 2 (m_2) มีค่าผิดพลาดร้อยละ q_1

   3.3 ให้จำนวนหน่วยตัวอย่างที่ถูกเลือกซ้ำในการเลือกครั้งที่ 3 (m_3) มีค่าผิดพลาดร้อยละ q_2

3.4 ให้จำนวนหน่วยตัวอย่างที่ถูกเลือกซ้ำในการเลือกครั้งที่ 2 และ 3 (m_{23}) มีค่าผิดพลาดร้อยละ q_3

3.5 ให้จำนวนหน่วยตัวอย่างที่ถูกเลือกซ้ำในการเลือกครั้งที่ 2 (m_2)และจำนวนหน่วยตัวอย่างที่ถูกเลือกซ้ำในการเลือกครั้งที่ 3 (m_3) มีค่าผิดพลาดร้อยละ q_1 และ q_2 ตามลำดับ

โดยที่ q_1 , q_2 และ q_3 มีค่าเป็น +10%, -10%, +20% และ -20%

หมายเหตุ : โดยตัวประมาณ Petersen จะทำการคำนวณตามกรณีที่ 3.1 และ 3.2 ตัวประมาณที่อาศัย Gibbs Sampler จะทำการคำนวณตามกรณีที่ 3.1, 3.2, 3.3 และ 3.5 ส่วนตัวประมาณที่พัฒนาจาก Petersen จะคำนวณทั้ง 5 กรณี ทั้งนี้ เพราะตัวประมาณ Petersen จะทำการเลือกตัวอย่างเพียงสองครั้ง ส่วนตัวประมาณที่อาศัย Gibbs Sampler จะพิจารณาเพียงว่าในแต่ละการเลือกตัวจะมีหน่วยตัวอย่างที่มีเครื่องหมายเป็นจำนวนเท่าใด โดยไม่ได้พิจารณาแยกด้วยว่ามีหน่วยตัวอย่างเป็นจำนวนเท่าใดที่ถูกเลือกซ้ำมากกว่าหนึ่งครั้ง (ไม่พิจารณา m_{23})

4. กำหนดค่าสัดส่วนระหว่างขนาดตัวอย่างในการเลือกครั้งที่ 2 และ 1 (n_2/n_1) มีค่าเป็น 0.50, 0.75 และ 1.00

5. ค่าพารามิเตอร์ (a,b) สำหรับการแจกแจง $\pi(p_i) \sim$ Beta(a,b) มีค่าเป็น(1,1)

เกณฑ์การตัดสินใจ

เนื่องจากในการศึกษาครั้งนี้ มีการกำหนดขนาดประชากรที่ต้องการประมาณ ทำให้ทราบขนาดประชากรที่แท้จริง ดังนั้นเกณฑ์ในการตัดสินใจว่าตัวประมาณใดที่สามารถประมาณขนาดประชากรได้ดีกว่ากันนั้น จะพิจารณาจากค่าความแตกต่างของค่าจริงและค่าประมาณที่ได้จากตัวประมาณทั้งสาม หากตัวประมาณใดมีค่าใกล้เคียงค่าจริงมากที่สุด จะยอมรับว่าตัวประมาณนั้นสามารถประมาณขนาดประชากรได้ดีกว่าตัวอื่น ๆ โดยใช้ค่าเฉลี่ยร้อยละของความคลาดเคลื่อนสัมพัทธ์สัมบูรณ์ (MAPE) ดังนี้

$$\text{MAPE} = \frac{1}{I}\sum_{i=1}^{I}\left(\frac{|N_i - \hat{N}_i|}{N_i}\right)\times 100$$

หอสมุดกลาง สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ประโยชน์ที่คาดว่าจะได้รับ

เพื่อเป็นแนวทางในการเลือกใช้ตัวประมาณที่เหมาะสมแก่ผู้วิจัยในการวิเคราะห์ปัญหาเกี่ยวกับการประมาณขนาดประชากร เมื่อทำการเลือกตัวอย่างแบบแคบเจอร์-รีแคบเจอร์