

รายงานการวิจัย

เรื่อง

การประมาณขนาดของประชากรโดยวิธี Capture – Recapture ในกรณีที่เป็

การสุ่มตัวอย่างแบบหลายครั้ง

Capture – Recapture Method for Estimation of Population Size on

Several Occasions

โดย

ผู้ช่วยศาสตราจารย์ นवलพรรณ หนังสือ

ผู้ช่วยศาสตราจารย์ ดร.วินัย โพธิ์สุวรรณ

งานวิจัยนี้ได้รับทุนสนับสนุนจากงบประมาณแผ่นดิน

สถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ

ปีงบประมาณ 2549

โครงการวิจัย	การประมาณขนาดประชากรโดยวิธี Capture – Recapture ในกรณีที่เป็นการสุ่มตัวอย่างแบบหลายครั้ง
ผู้ทำการวิจัย	1. ผู้ช่วยศาสตราจารย์ นวลพรรณ หนังสือ 2. ผู้ช่วยศาสตราจารย์ ดร. วินัย โพธิ์สุวรรณ ภาควิชาสถิติประยุกต์ คณะวิทยาศาสตร์ประยุกต์ สถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ
ได้รับทุนอุดหนุน	ทุนอุดหนุนงบประมาณแผ่นดิน ประจำปี 2549

บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อนำเสนอวิธีการประมาณขนาดของประชากรโดยวิธี Capture – Recapture กรณีที่เป็นการสุ่มตัวอย่างแบบหลายครั้งโดยอาศัยแนวคิดที่พัฒนามาจากวิธีของ Jackson (Modified Jackson Method) และเพื่อเปรียบเทียบวิธีการประมาณดังกล่าวกับวิธีเดิม รวมทั้งเพื่อประมาณค่าช่วงความเชื่อมั่นของจำนวนของสมาชิกที่ถูกทำเครื่องหมายและสุ่มได้ในการสุ่มตัวอย่างครั้งถัดไปที่เหมาะสม ในการวิจัยครั้งนี้ได้พัฒนาวิธีการประมาณขนาดของประชากรขึ้นใหม่ 2 วิธีคือ วิธีที่พัฒนามาจาก Jackson โดยใช้ค่ามัธยฐานของเวลา (Modified Jackson using Median Time Method) และวิธีที่พัฒนามาจาก Jackson โดยใช้ค่ามัธยฐานของจำนวนของสมาชิกที่ถูกทำเครื่องหมายและสุ่มได้ในการสุ่มตัวอย่างครั้งถัดไป (Modified Jackson using Median m_i Method) โดยพิจารณาจากค่าเฉลี่ยของเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์เป็นเกณฑ์ในการเปรียบเทียบค่าประมาณจากวิธีการดังกล่าวข้างต้น ทั้ง 3 วิธี ซึ่งได้ทำการจำลองค่าประชากร $N = 500$ และ $1,000$ โดยแบ่งออกเป็นกรณีที่มี M และ n มีค่าเท่ากันและมีค่าแตกต่างกัน โดยในแต่ละกรณีทำซ้ำ 10,000 ครั้ง โดยใช้โปรแกรม Minitab Macros จากการเปรียบเทียบตัวแบบพยากรณ์ที่ได้จากวิธีการต่าง ๆ ทั้ง 3 วิธี โดยพิจารณาจากค่าเฉลี่ยของเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์ (MAPE) แล้ว พบว่าเมื่อพิจารณาค่า MAPE ในทุกกรณี พบว่าวิธีที่พัฒนามาจาก Jackson โดยใช้ค่ามัธยฐานของเวลา จะให้ค่า MAPE ต่ำที่สุด รองลงมาคือวิธีที่พัฒนามาจาก Jackson โดยใช้ค่ามัธยฐานของจำนวนของสมาชิกที่ถูกทำเครื่องหมายและสุ่มได้ในการสุ่มตัวอย่างครั้งถัดไป และวิธีที่พัฒนามาจาก Jackson ตามลำดับ และเมื่อพิจารณาช่วงความเชื่อมั่น 95% ของจำนวนของสมาชิกที่ถูกทำเครื่องหมายและสุ่มได้ในการสุ่มตัวอย่างครั้งถัดไป พบว่าถ้าสัดส่วนการสุ่มตัวอย่างเพิ่มขึ้นช่วงความเชื่อมั่นของจำนวนสมาชิกที่ถูกทำเครื่องหมายและสุ่มได้ในการสุ่มตัวอย่างครั้งถัดไปจะเข้าใกล้ค่าของจำนวนของสมาชิกที่ทำการสุ่มได้และถูกทำเครื่องหมายไว้ในการสุ่มตัวอย่างครั้งแรก (M) แต่จะมีค่าใกล้เคียงกันในแต่ละกรณีที่มีสัดส่วนการสุ่มตัวอย่างเท่ากันและจำนวนครั้งของการสุ่มตัวอย่างที่แตกต่างกัน

**RESEARCH NAME :CAPTURE – RECAPTURE METHOD FOR ESTIMATION OF
POPULATION SIZE ON SEVERAL OCCASIONS.**

**RESEARCHER : 1. ASSISTANT PROFESSOR NUANPAN NANGSUE
2. ASSISTANT PROFESSOR DR. WINAI BODHISUWAN
APPLIED STATISTICAL DEPARTMENT
FACULTY OF APPLIED SCIENCE
KING MONGKUT'S INSTITUTE OF TECHNOLOGY NORTH
BANGKOK**

YEAR : 2006

Abstract

This research proposed two new methods for the estimation of population sizes on several occasions using capture-recapture methods for an open population. These methods were developed by modifying the Jackson classical method. We also compared the new method with the old one and estimated the confidence interval for m_i , the number of tagged individuals captured on occasion i . In this research, we proposed two methods. In the first method namely modified Jackson using median time, we estimated the population size at the median time to replace at time zero. In the second one namely modified Jackson using median m_i , we built the regression model between the estimate of population size on each occasion and m_i , the number of tagged individuals captured on occasion i , because their relationship is an inverse relationship and the population size was estimated from the median m_i . Three methods were compared in a simulation study using population sizes (N) of 500 and 1,000 for different sample sizes M , the number of individuals marked in the first sample and n , total number of individuals captured in the second sample. In each case we did 10,000 replications using Minitab Macros.

Comparison of the estimation from the three methods showed that the modified Jackson using median time method gave the least mean absolute percentage error and the modified Jackson was the worst. For the 95% confidence intervals for m_i , if the sampling fraction increased then the confidence interval for m_i was close to M . It was nearly the same for the same sampling fraction and different numbers of occasions.

กิตติกรรมประกาศ

งานวิจัยนี้สำเร็จลุล่วงได้ด้วยดี ทางคณะผู้วิจัยขอกราบขอบพระคุณครู อาจารย์ทุกท่านที่ได้
ประสิทธิประสาทวิชาความรู้ให้แก่คณะผู้วิจัย และขอขอบคุณแหล่งทุนวิจัยที่ให้ทุนสนับสนุนในการ
ทำวิจัยในครั้งนี้

สุดท้ายนี้ ผู้วิจัยขอกราบขอบพระคุณคุณพ่อคุณแม่ และครอบครัวที่ให้ความห่วงใยและเป็น
กำลังใจให้กับคณะผู้วิจัยเสมอมา

คณะผู้วิจัย

สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	ก
บทคัดย่อภาษาอังกฤษ.....	ข
กิตติกรรมประกาศ.....	ค
สารบัญ.....	ง
สารบัญตาราง.....	จ
สารบัญรูป.....	ฉ
บทที่ 1 บทนำ.....	1
1.1 ความเป็นมาและความสำคัญของปัญหา.....	1
1.2 วัตถุประสงค์ของการวิจัย.....	3
1.3 ขอบเขตของการวิจัย.....	3
1.4 เกณฑ์ที่ใช้ในการพิจารณา.....	3
1.5 ประโยชน์ที่คาดว่าจะได้รับ.....	4
บทที่ 2 ทฤษฎีที่เกี่ยวข้อง.....	5
2.1 งานวิจัยที่เกี่ยวข้อง.....	5
2.2 ทฤษฎีสถิติที่เกี่ยวข้องกับงานวิจัย.....	6
บทที่ 3 วิธีดำเนินงานวิจัย.....	12
บทที่ 4 ผลการวิเคราะห์ข้อมูล.....	15
บทที่ 5 สรุปผลและข้อเสนอแนะ.....	30
5.1 สรุปผลการวิเคราะห์ข้อมูล.....	30
5.2 ข้อเสนอแนะ.....	31
บรรณานุกรม.....	32
ภาคผนวก	34

สารบัญรูป (ต่อ)

รูปที่	หน้า
4.12 ค่าเฉลี่ยของเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์ (MAPE) ของวิธีการต่าง ๆ สำหรับ จำนวนครั้งของการสุ่มที่แตกต่างกัน กรณีประชากรขนาด $N = 1,000$ และ $M = n = 800$	24
4.13 ค่าเฉลี่ยของเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์ (MAPE) ของวิธีการต่าง ๆ สำหรับ จำนวนครั้งของการสุ่มที่แตกต่างกัน กรณีประชากรขนาด $N = 1,000$ และ $M = 100$, $n = 200$	25
4.14 ค่าเฉลี่ยของเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์ (MAPE) ของวิธีการต่าง ๆ สำหรับ จำนวนครั้งของการสุ่มที่แตกต่างกัน กรณีประชากรขนาด $N = 1,000$ และ $M = 200$, $n = 100$	26
4.15 ค่าเฉลี่ยของเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์ (MAPE) ของวิธีการต่าง ๆ สำหรับ จำนวนครั้งของการสุ่มที่แตกต่างกัน กรณีประชากรขนาด $N = 1,000$ และ $M = 200$, $n = 400$	26
4.16 ค่าเฉลี่ยของเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์ (MAPE) ของวิธีการต่าง ๆ สำหรับ จำนวนครั้งของการสุ่มที่แตกต่างกัน กรณีประชากรขนาด $N = 1,000$ และ $M = 400$, $n = 200$	27

สารบัญตาราง

ตารางที่	หน้า
4.1 ค่า MAPE สำหรับจำนวนครั้งของการสุ่มที่แตกต่างกัน กรณีประชากรขนาด N = 500 และขนาดตัวอย่างเหมือนกัน.....	15
4.2 ค่า MAPE สำหรับจำนวนครั้งของการสุ่มที่แตกต่างกัน กรณีประชากรขนาด N = 500 และขนาดตัวอย่างแตกต่างกัน.....	18
4.3 ค่า MAPE สำหรับจำนวนครั้งของการสุ่มที่แตกต่างกัน กรณีประชากรขนาด N = 1,000 และขนาดตัวอย่างเหมือนกัน.....	21
4.4 ค่า MAPE สำหรับจำนวนครั้งของการสุ่มที่แตกต่างกัน กรณีประชากรขนาด N = 1,000 และขนาดตัวอย่างแตกต่างกัน.....	24
4.5 ช่วงความเชื่อมั่น 95% ของจำนวนของสมาชิกที่ถูกทำเครื่องหมายและสุ่มได้ในการ สุ่มตัวอย่างครั้งถัดไป กรณีประชากรขนาด N = 500 และขนาดตัวอย่างเหมือนกัน.....	27
4.6 ช่วงความเชื่อมั่น 95% ของจำนวนของสมาชิกที่ถูกทำเครื่องหมายและสุ่มได้ในการ สุ่มตัวอย่างครั้งถัดไป กรณีประชากรขนาด N = 500 และขนาดตัวอย่างแตกต่างกัน.....	28
4.7 ช่วงความเชื่อมั่น 95% ของจำนวนของสมาชิกที่ถูกทำเครื่องหมายและสุ่มได้ในการ สุ่มตัวอย่างครั้งถัดไป กรณีประชากรขนาด N = 1,000 และขนาดตัวอย่างเหมือนกัน....	28
4.8 ช่วงความเชื่อมั่น 95% ของจำนวนของสมาชิกที่ถูกทำเครื่องหมายและสุ่มได้ในการ สุ่มตัวอย่างครั้งถัดไป กรณีประชากรขนาด N = 1,000 และขนาดตัวอย่างต่างกัน.....	29

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

การประมาณขนาดของประชากรโดยวิธี Capture - Recapture นั้น โดยทั่วไปมักจะนำมาใช้ในการประมาณขนาดของประชากรสัตว์ซึ่งมีอยู่จำนวนมาก เช่น การประมาณจำนวนปลาที่อยู่ในบ่อน้ำ จำนวนกวางที่อยู่ในป่า เป็นต้น เนื่องจากการเก็บรวบรวมข้อมูลจากทุกหน่วยของประชากรทำได้ยากหรือบางครั้งไม่สามารถทำได้เพราะประชากรสัตว์มีการเคลื่อนที่อยู่ตลอดเวลา แต่ในปัจจุบันพบว่าวิธีการนี้นิยมนำมาใช้ในศาสตร์สาขาวิชาอื่น ๆ เช่น สาขาการแพทย์ที่เกี่ยวข้องกับโรคระบาด ด้านสังคมศาสตร์ และทางด้านอุตสาหกรรมซึ่งเป็นปัญหาที่มีความสำคัญเป็นอย่างมาก เช่น Aaron และคณะ (2003) ได้ทำการประมาณขนาดประชากรที่อยู่ในกลุ่มรักร่วมเพศหญิง (Lesbian) โดยใช้วิธีการ Capture-Recapture Mastro และคณะ (1994) ได้นำวิธีการ Capture - Recapture มาใช้ในการประมาณจำนวนผู้ติดยาเสพติดซึ่งได้รับเชื้อ HIV ในประเทศไทย Fisher และคณะ (1994) ทำการประมาณจำนวนของผู้ที่ไร้ที่อยู่อาศัยและผู้ไร้ที่อยู่อาศัยซึ่งมีอาการป่วยทางจิต ทางตะวันออกเฉียงเหนือของ Westminster โดยใช้การวิเคราะห์ Capture-Recapture Pezzotti และคณะ (2003) นำมาใช้ในการประมาณจำนวนของคนที่ตรวจพบว่าติดเชื้อ HIV โดยใช้ข้อมูล Cross-Linking จาก 4 แหล่งข้อมูล และ Sartor และ Walckiers (2001) นำมาใช้ในการประมาณจำนวนของปัญหาที่เกิดขึ้นในการใช้ยาเสพติดในเบลเยียม เป็นต้น

การประมาณโดยวิธีนี้จะทำการสุ่มตัวอย่างขนาด M จากประชากรที่มีขนาด N และนำตัวอย่างที่สุ่มได้มาทำเครื่องหมายไว้ แล้วปล่อยกลับไปสู่ประชากร รอจนช่วงเวลาหนึ่งเพื่อให้ตัวอย่างสุ่มที่ถูกทำเครื่องหมายรวมอยู่ในประชากร หลังจากนั้นจะทำการสุ่มตัวอย่างอีกครั้งหนึ่งขนาด n แล้วทำการจดบันทึกตัวอย่างสุ่มที่ถูกทำเครื่องหมายไว้ ซึ่งการสุ่มลักษณะนี้เป็นการสุ่มตัวอย่างแบบ 2 ครั้ง หรือจะทำการสุ่มตัวอย่างในครั้งต่อไปก็ได้

วิธีการ Capture - Recapture ที่รู้จักกันอย่างแพร่หลายคือวิธีของ Lincoln-Petersen (1896,1930) ซึ่งเป็นวิธีการประมาณขนาดของประชากรอย่างง่าย โดยเป็นการสุ่มตัวอย่างแบบ 2 ครั้ง แต่มีข้อจำกัดว่าประชากรที่ศึกษาเป็นประชากรแบบปิด (Close Population) นั่นคือในทางสถิติประชากรไม่มีการเกิดหรือตาย และทางภูมิศาสตร์ไม่มีการอพยพย้ายถิ่นหรือการอพยพเข้ามาอยู่ในถิ่นใหม่ หรือในกรณีที่เป็นการสุ่มตัวอย่างแบบหลายครั้งได้แก่วิธีการของ Schnabel (1938)

แต่เนื่องจากในความเป็นจริงประชากรอาจมีการเกิดหรือตาย หรืออพยพย้ายถิ่นหรือการอพยพเข้ามาอยู่ในถิ่นใหม่ซึ่งเรียกว่าเป็นประชากรแบบเปิด (Open Population) ทำให้ต่อมาจึงได้มีผู้พัฒนาวิธีการที่ใช้ในการประมาณขนาดของประชากรในกรณีนี้ซึ่งมีด้วยกันหลายวิธี เช่นวิธีของ Jackson (1939,1948) ซึ่งนำมาใช้ในการประมาณแมลงเซทซีในประเทศแอฟริกา หรือวิธีของ Jolly Seber (1965,1982) ซึ่งนิยมใช้กันอย่างแพร่หลาย โดยวิธีการเหล่านี้เป็นการสุ่มตัวอย่างแบบหลายครั้ง (Several Occasions) ผู้วิจัยได้ทำการศึกษาโดยอาศัยแนวคิดที่พัฒนามาจากวิธีของ Jackson เนื่องจากเป็นวิธีการที่สะดวกและไม่ยุ่งยากซึ่งจะทำการประมาณขนาดของประชากรโดยทำการสุ่มตัวอย่างในครั้งแรกและทำเครื่องหมายไว้ (M_1) หลังจากนั้นปล่อยกลับไปสู่ประชากร รอจนช่วงเวลาหนึ่งแล้วจึงทำการสุ่มตัวอย่างครั้งต่อ ๆ มา ณ ช่วงเวลาต่าง ๆ กัน โดยเมื่อสุ่มตัวอย่างได้ในแต่ละครั้งแล้วให้นับจำนวนสมาชิกที่ถูกทำเครื่องหมายไว้ในครั้งแรกและสุ่มได้ในครั้งนั้น (m_i) แล้วจึงปล่อยคืนกลับไปสู่ประชากรอีกครั้งหนึ่ง เนื่องจากประชากรเป็นประชากรแบบเปิด ดังนั้นจะพบว่าสัดส่วนของจำนวนสมาชิกที่ถูกทำเครื่องหมายและจับได้ในครั้งต่อ ๆ มาจะมีจำนวนลดลงเมื่อเวลาเปลี่ยนแปลงไป เพราะวิธีการนี้จะทำเครื่องหมายให้กับสมาชิกแต่ละตัวที่สุ่มได้ในการสุ่มตัวอย่างครั้งแรกเท่านั้น โดยวิธีการที่พัฒนามาจากวิธีของ Jackson จะทำการประมาณขนาดของประชากรโดยอาศัยเทคนิคของการวิเคราะห์การถดถอยเพื่อสร้างสมการแสดงความสัมพันธ์ระหว่างขนาดของประชากรที่ประมาณได้โดยใช้วิธีการของ Lincoln – Petersen ประมาณขนาดประชากรสำหรับการสุ่มตัวอย่างในแต่ละครั้งกับตัวแปรเวลา แล้วจึงประมาณค่าขนาดของประชากรที่เวลาเริ่มต้นซึ่งได้ทำการสุ่มตัวอย่างและทำเครื่องหมายไว้ในครั้งแรก นั่นคือจะได้ว่าระยะตัดแกน Y เป็นค่าประมาณขนาดของประชากร แต่ผู้วิจัยพบว่าในทางปฏิบัติเมื่อเวลาผ่านไปเป็นเวลานานจำนวนสมาชิกที่ถูกทำเครื่องหมายไว้ในครั้งแรกและถูกจับได้ในครั้งต่อ ๆ มาจะมีจำนวนลดลงเรื่อยๆ และเมื่อประมาณค่าขนาดของประชากรโดยวิธี Lincoln – Petersen สำหรับการสุ่มตัวอย่างแต่ละครั้งจะทำให้ขนาดของประชากรที่ประมาณได้มีค่าสูงกว่าความเป็นจริง แต่ในทางตรงข้ามถ้าทำการประมาณขนาดของประชากร ณ ตำแหน่งของเวลาที่ทำการสุ่มตัวอย่างในครั้งแรกๆ ซึ่งมีแนวโน้มว่าจำนวนสมาชิกที่สุ่มได้และถูกทำเครื่องหมายไว้ในครั้งแรกๆ จะมีความชันของเส้นถดถอยที่ได้มีค่ามากเป็นผลให้ระยะตัดแกน Y มีค่าน้อย ดังนั้นเมื่อประมาณค่าขนาดของประชากรที่เวลาเริ่มต้นจึงอาจได้ค่าประมาณขนาดของประชากรที่ต่ำกว่าความเป็นจริงมาก ผู้วิจัยจึงได้เล็งเห็นถึงปัญหาในข้อนี้และได้อาศัยแนวคิดของวิธีการนี้นำมาพัฒนาวิธีการที่ใช้ในการประมาณขนาดของประชากรโดยวิธี Capture – Recapture ในกรณีที่เป็นประชากรแบบเปิดและเป็นการสุ่มตัวอย่างแบบหลายครั้งขึ้นใหม่ เพื่อให้ได้ค่าประมาณขนาดของประชากรที่มีความคลาดเคลื่อนต่ำที่สุด

1.2 วัตถุประสงค์ของโครงการวิจัย

1. เพื่อนำเสนอวิธีการประมาณขนาดของประชากรโดยวิธี Capture – Recapture กรณีที่เป็นการสุ่มตัวอย่างแบบหลายครั้งขึ้นใหม่โดยอาศัยแนวคิดที่พัฒนามาจากวิธีของ Jackson
2. เพื่อเปรียบเทียบวิธีการประมาณขนาดของประชากรโดยวิธี Capture – Recapture กรณีที่เป็นการสุ่มตัวอย่างแบบหลายครั้งที่ผู้วิจัยได้พัฒนาขึ้นกับวิธีการที่พัฒนามาจากวิธีของ Jackson
3. เพื่อประมาณค่าช่วงความเชื่อมั่นของจำนวนของสมาชิกที่ถูกทำเครื่องหมายและสุ่มได้ในการสุ่มตัวอย่างครั้งถัดไปที่เหมาะสม

1.3 ขอบเขตของการวิจัย

การวิจัยนี้มีขอบเขตของโครงการวิจัยดังต่อไปนี้

1. ในงานวิจัยนี้จะศึกษาการประมาณขนาดของประชากรโดยวิธี Capture – Recapture สำหรับประชากรแบบเปิด เฉพาะกรณีที่เป็นการสุ่มตัวอย่างแบบหลายครั้ง ซึ่งทำเครื่องหมายให้กับสมาชิกแต่ละตัวที่สุ่มได้ในการสุ่มตัวอย่างครั้งแรกเท่านั้น โดยมีวิธีการที่ศึกษา 3 วิธีดังนี้คือ วิธีที่พัฒนามาจาก Jackson (Modified Jackson Method) และวิธีการประมาณขนาดของประชากรที่พัฒนาขึ้นมาใหม่โดยอาศัยแนวคิดของ Jackson คือวิธีที่พัฒนามาจาก Jackson โดยใช้ค่ามัธยฐานของเวลา (Modified Jackson using Median Time Method) และวิธีที่พัฒนามาจาก Jackson โดยใช้ค่ามัธยฐานของจำนวนของสมาชิกที่ถูกทำเครื่องหมายและสุ่มได้ในการสุ่มตัวอย่างครั้งถัดไป (Modified Jackson using Median m_i Method)
2. การศึกษานี้จะกำหนดค่า $N = 500$ และ $1,000$ โดยแบ่งออกเป็นกรณีที่ M และ n มีค่าเท่ากันและมีค่าแตกต่างกัน

1.4 เกณฑ์ที่ใช้ในการพิจารณา

เกณฑ์ที่ใช้ในการวัดความถูกต้องของการพยากรณ์ในการวิจัยครั้งนี้คือค่าเฉลี่ยของเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์ (Mean Absolute Percentage Error) ซึ่งนำมาใช้ในการเปรียบเทียบค่าประมาณจากวิธีการดังกล่าวข้างต้น ทั้ง 3 วิธี ว่าวิธีการใดเหมาะสมที่สุด

$$MAPE = (100/n) \sum_{i=1}^n \left| \frac{e_i}{Y_i} \right|$$

โดย e_t คือความคลาดเคลื่อนของการพยากรณ์

$$\text{ซึ่งคำนวณได้ดังนี้ } e_t = Y_t - \hat{Y}_t$$

Y_t คือค่าจริงของตัวแปรตาม

$\hat{Y}_t$ คือค่าพยากรณ์ของตัวแปรตาม

n คือจำนวนข้อมูล

โดยพิจารณาจากค่า MAPE ของการพยากรณ์ในแต่ละวิธี และจะเลือกตัวแบบที่มีค่า MAPE น้อยสุดเป็นตัวแบบพยากรณ์

1.5 ประโยชน์ที่คาดว่าจะได้รับ

1. ได้วิธีการประมาณขนาดของประชากร โดยวิธี Capture – Recapture สำหรับประชากรแบบเปิดกรณีที่เป็นการสุ่มตัวอย่างแบบหลายครั้งขึ้นใหม่

2. หน่วยงานที่เกี่ยวข้องเช่นหน่วยงานของรัฐบาลด้านสาธารณสุขที่เกี่ยวกับโรคระบาด หน่วยงานทางด้านสังคมศาสตร์และทางด้านอุตสาหกรรม สามารถนำเทคนิควิธีนี้ไปใช้ประโยชน์ในการประมาณขนาดของประชากรที่ไม่ทราบค่าได้ ซึ่งวิธีที่พัฒนาขึ้นเป็นวิธีการที่สะดวกและทำการประมวลผลได้ง่าย

บทที่ 2

ทฤษฎีที่เกี่ยวข้อง

2.1 งานวิจัยที่เกี่ยวข้อง

Lincoln-Petersen (1896,1930) ได้เสนอวิธีการประมาณขนาดของประชากร โดยวิธีการ Capture - Recapture ซึ่งเป็นที่รู้จักกันอย่างแพร่หลาย โดยอยู่ภายใต้ข้อตกลงที่ว่าประชากรที่ศึกษาเป็นประชากรแบบเปิด นั่นคือในทางสถิติประชากรไม่มีการเกิดหรือตาย และทางภูมิศาสตร์ไม่มีการอพยพย้ายถิ่นหรือการอพยพเข้ามาอยู่ในถิ่นใหม่ สมาชิกแต่ละตัวมีโอกาสถูกเลือกเป็นตัวอย่างด้วยความน่าจะเป็นเท่า ๆ กัน และความน่าจะเป็นนี้ไม่เปลี่ยนแปลงตามเวลา การสุ่มตัวอย่างในแต่ละครั้งเป็นไปอย่างสุ่ม และสมาชิกที่ถูกทำเครื่องหมายจะไม่สูญหายในระหว่างการสุ่มทั้งสองครั้ง รวมทั้งถูกบันทึกและรายงานการค้นพบอีกครั้งในการสุ่มตัวอย่างครั้งที่ 2 โดยวิธีของ Lincoln -Petersen เป็นวิธีการประมาณขนาดของประชากรอย่างง่ายที่อยู่ภายใต้พื้นฐานว่าสัดส่วนของสมาชิกที่ถูกทำเครื่องหมายในการสุ่มตัวอย่างครั้งที่ 2 ควรจะเท่ากับสัดส่วนของสมาชิกที่ถูกทำเครื่องหมายในประชากร แต่ตัวประมาณของ Lincoln-Petersen เป็นตัวประมาณที่เอนเอียง (Bias Estimator) และถ้าจำนวนของสมาชิกที่ถูกทำเครื่องหมาย และสุ่มได้ในการสุ่มตัวอย่างครั้งที่ 2 (m) มีขนาดเล็กอาจทำให้ค่าประมาณที่ได้มีค่าสูงกว่าความเป็นจริง รวมทั้งถ้า $m = 0$ จะทำให้ไม่สามารถหาค่าประมาณได้

ต่อมา Chapman (1951) ได้เสนอตัวประมาณซึ่งเป็นตัวประมาณที่ไม่เอนเอียง โดยที่ตัวประมาณของ Chapman อยู่ภายใต้ข้อตกลงเพิ่มเติมที่ว่า การสุ่มตัวอย่างในครั้งที่ 2 เป็นการสุ่มแบบไม่ใส่คืน และจำนวนสมาชิกที่ถูกทำเครื่องหมายและสุ่มได้อีกในครั้งที่ 2 (m) มีการแจกแจงแบบไฮเปอร์จีโอเมตริก ต่อมา Seber (1982) ได้เสนอว่าตัวประมาณที่ได้จากวิธีการนี้จะเป็นตัวประมาณที่ไม่เอนเอียงถ้า $(n+M) \geq N$ และจะใกล้เคียงกับตัวประมาณไม่เอนเอียงถ้า $m > 7$

Jackson (1939,1948) ได้เสนอวิธีการเพื่อนำมาใช้ในการประมาณแมลงเซตซีในประเทศแอฟริกา ซึ่งทำการประมาณขนาดของประชากรโดยทำการสุ่มตัวอย่างในครั้งแรกและทำเครื่องหมายไว้ (M_1) หลังจากนั้นปล่อยกลับไปสู่ประชากร รอจนช่วงเวลาหนึ่งแล้วจึงทำการสุ่มตัวอย่างครั้งต่อ ๆ มา ณ ช่วงเวลาต่าง ๆ ที่เปลี่ยนแปลงไป โดยอาศัยพื้นฐานที่ว่าสัดส่วนของจำนวนสมาชิกที่ถูกทำเครื่องหมายและจับได้ในครั้งต่อ ๆ มาจะมีจำนวนลดลงเมื่อเวลาเปลี่ยนแปลงไป ซึ่งจากข้อสมมตินี้จะทำให้เวลาที่มีความสัมพันธ์แบบแปรผันตามกับขนาดของประชากรที่ประมาณขึ้น

Jackson ใน

การทำการประมาณขนาดของประชากรโดยอาศัยเทคนิคของการวิเคราะห์การถดถอยเพื่อสร้างสมการแสดงความสัมพันธ์ระหว่างขนาดของประชากรที่ประมาณขึ้นโดยใช้วิธีการของ Lincoln – Petersen ในการสุ่มแต่ละครั้งกับตัวแปรเวลา แล้วจึงประมาณค่าขนาดของประชากรที่เวลาเริ่มต้น ซึ่งได้ทำการสุ่มตัวอย่างในครั้งแรกและทำเครื่องหมายไว้

2.2 ทฤษฎีสถิติที่เกี่ยวข้องกับงานวิจัย

การประมาณขนาดของประชากรโดยวิธีการ Capture-Recapture ที่รู้จักกันอย่างแพร่หลายคือวิธีของ Lincoln-Petersen (1896,1930) ซึ่งเป็นวิธีการประมาณขนาดของประชากรอย่างง่าย โดยอยู่ภายใต้ข้อตกลงที่ว่าประชากรที่ศึกษาเป็นประชากรแบบเปิด นั่นคือในทางสถิติประชากรไม่มีการเกิดหรือตาย และทางภูมิศาสตร์ไม่มีการอพยพย้ายถิ่นหรือการอพยพเข้ามาอยู่ในถิ่นใหม่ สมาชิกแต่ละตัวมีโอกาสถูกเลือกเป็นตัวอย่างด้วยความน่าจะเป็นเท่า ๆ กัน และความน่าจะเป็นนี้ไม่เปลี่ยนแปลงตามเวลา การสุ่มตัวอย่างในแต่ละครั้งเป็นไปอย่างสุ่ม และสมาชิกที่ถูกทำเครื่องหมายจะไม่สูญหายในระหว่างการสุ่มทั้งสองครั้ง รวมทั้งถูกบันทึกและรายงานการค้นพบอีกครั้งในการสุ่มตัวอย่างครั้งที่ 2

วิธีของ Lincoln-Petersen เป็นวิธีการประมาณขนาดของประชากรอย่างง่ายที่อยู่ภายใต้พื้นฐานว่าสัดส่วนของสมาชิกที่ถูกทำเครื่องหมายในการสุ่มตัวอย่างครั้งที่ 2 ควรจะเท่ากับสัดส่วนของสมาชิกที่ถูกทำเครื่องหมายในประชากร

$$\frac{m}{n} = \frac{M}{N}$$

ดังนั้นจะได้ว่า

$$\hat{N}_p = \frac{n M}{m} \quad (1)$$

สัญลักษณ์ที่ใช้

M = จำนวนของสมาชิกที่ทำการสุ่มได้และถูกทำเครื่องหมายไว้ใน การสุ่มตัวอย่างครั้งแรก

n = จำนวนของสมาชิกที่ทำการสุ่ม ในการสุ่มตัวอย่างครั้งที่ 2

m = จำนวนของสมาชิกที่ถูกทำเครื่องหมาย และสุ่มได้ในการสุ่มตัวอย่างครั้งที่ 2

N = ขนาดของประชากรที่ประมาณ

อีกวิธีการหนึ่งที่ใช้ในการหาค่า $\hat{N}$ ซึ่งเป็นตัวประมาณ MLE (Maximum Likelihood Estimator) โดยใช้การแจกแจงแบบไฮเปอร์จีอเมตริก (Hypergeometric Distribution) ซึ่งเป็นการสุ่มแบบไม่ใส่คืน (Sampling Without Replacement)

$$P(m = m) = \frac{\binom{M}{m} \binom{N-M}{n-m}}{\binom{N}{n}}$$

และได้ Likelihood เป็น

$$L(N; M, n, m) = \frac{\binom{M}{m} \binom{N-M}{n-m}}{\binom{N}{n}}$$

จากนั้นทำการหาค่า MLE จะได้ $\hat{N}$ เช่นเดียวกับสมการที่ (1) หรือจะหาตัวประมาณโดยใช้การแจกแจงทวินาม (Binomial Distribution) ประมาณการแจกแจงแบบไฮเปอร์จีอเมตริกในกรณีที่สุ่มตัวอย่างขนาดใหญ่ หรือในกรณีที่เป็นการสุ่มแบบใส่คืน (Sampling with Replacement) โดยกำหนดให้สถานภาพที่ถูกทำเครื่องหมายหรือไม่ถูกทำเครื่องหมายของสมาชิกแต่ละตัวในการสุ่มตัวอย่างครั้งที่ 2 เป็นการทดลองแบบเบอร์นูลลี (Benoulli Experiment) นั่นคือสมาชิกแต่ละตัวที่ถูกสุ่มขึ้นมา มีความน่าจะเป็นที่จะถูกทำเครื่องหมายดังต่อไปนี้

$$p = \frac{M}{N}$$

กำหนดให้

$$y_i = \begin{cases} 0 & , \text{ ถ้าไม่ถูกทำเครื่องหมาย} \\ 1 & , \text{ ถ้าถูกทำเครื่องหมาย} \end{cases} \quad i = 1, 2, \dots, n$$

โดยกำหนดให้จำนวนที่ถูกทำเครื่องหมายในการสุ่มตัวอย่างครั้งที่ 2 เป็น

$$m = \sum_{i=1}^n y_i$$

ตัวแปรสุ่ม m มีการแจกแจงแบบทวินามด้วยพารามิเตอร์ n และ p โดยที่ทราบค่า n แต่ไม่ทราบค่า p (เนื่องจากไม่ทราบค่า N)

ความน่าจะเป็นของค่าสังเกต $m = m$ ของสมาชิกแต่ละตัวที่ถูกทำเครื่องหมายและจับได้ในการสุ่มตัวอย่างครั้งที่ 2 เป็น

$$\begin{aligned}
L(N:M,n,m) &= P(m = m/p, n) \\
&= \binom{n}{m} p^m (1-p)^{n-m} \\
&= \binom{n}{m} \left(\frac{M}{N}\right)^m \left(1 - \frac{M}{N}\right)^{n-m}
\end{aligned}$$

และทำการหาค่า MLE จะได้ $\hat{N}$ เช่นเดียวกัน

ส่วนการประมาณค่าความแปรปรวนของ ตัวประมาณ Lincoln-Petersen มีสูตรดังนี้

$$\text{var}(\hat{N}_p) = \frac{Mn(M-m)(n-m)}{m^3} \quad (2)$$

ตัวประมาณของ Lincoln-Petersen เป็นตัวประมาณที่เอนเอียง (Bias Estimator) และถ้า m มีขนาดเล็กอาจทำให้ค่าประมาณที่ได้มีค่าสูงกว่าความเป็นจริง รวมทั้งถ้า $m = 0$ จะทำให้ไม่สามารถหาค่าประมาณได้

จากปัญหาที่กล่าวไว้สำหรับตัวประมาณของ Lincoln-Petersen ต่อมา Chapman (1951) ได้เสนอตัวประมาณที่เป็นตัวประมาณที่ไม่เอนเอียงดังนี้

$$\hat{N}_c = \frac{(n+1)(M+1)}{(m+1)} - 1 \quad (3)$$

ซึ่งมีค่าประมาณของความแปรปรวนดังนี้

$$\text{var}(\hat{N}_c) = \frac{(M+1)(n+1)(M-m)(n-m)}{(m+1)^2(m+2)} \quad (4)$$

โดยที่ตัวประมาณของ Chapman อยู่ภายใต้ข้อตกลงเพิ่มเติมที่ว่า การสุ่มตัวอย่างในครั้งที่ 2 เป็นการสุ่มแบบไม่ใส่คืน และจำนวนสมาชิกที่ถูกทำเครื่องหมายและสุ่มได้อีกในครั้งที่ 2 (m) มีการแจกแจงแบบไฮเปอร์จีโอเมตริก ต่อมา Seber (1982) ได้เสนอว่าตัวประมาณที่ได้จากวิธีการนี้จะเป็นตัวประมาณที่ไม่เอนเอียงถ้า $(n+M) \geq N$ และจะใกล้เคียงกับตัวประมาณไม่เอนเอียงถ้า $m > 7$

จากที่กล่าวมาข้างต้นเป็นวิธีการประมาณขนาดของประชากรในกรณีที่ประชากรเป็นแบบปิด แต่จะพบว่าในความเป็นจริงประชากรอาจจะมีลักษณะเป็นประชากรแบบเปิดนั่นคือในทางสถิติ ประชากรอาจมีการเกิดหรือตาย และทางภูมิศาสตร์อาจมีการอพยพย้ายถิ่นหรือการอพยพเข้ามาอยู่ในถิ่นใหม่ จึงได้มีผู้พัฒนาการประมาณขนาดของประชากรโดยวิธี Capture – Recapture

Jackson (1939,1948) ได้เสนอวิธีการเพื่อนำมาใช้ในการประมาณแมลงเซทซีในประเทศแอฟริกา ซึ่งทำการประมาณขนาดของประชากรโดยทำการสุ่มตัวอย่างในครั้งแรกและทำเครื่องหมายไว้ (M_1) หลังจากนั้นปล่อยกลับไปสู่ประชากร รอจนช่วงเวลาหนึ่งแล้วจึงทำการสุ่มตัวอย่างครั้งต่อ ๆ มา ณ ช่วงเวลาต่าง ๆ ที่เปลี่ยนแปลงไป โดยอาศัยพื้นฐานที่ว่าสัดส่วนของจำนวนสมาชิกที่ถูกทำเครื่องหมายและจับได้ในครั้งต่อ ๆ มาจะมีจำนวนลดลงเมื่อเวลาเปลี่ยนแปลงไป ซึ่งจากข้อสมมตินี้จะทำให้เวลาที่มีความสัมพันธ์แบบแปรผันตามกับขนาดของประชากรที่ประมาณขึ้นในแต่ละครั้งที่ทำการสุ่มตัวอย่างในแต่ละช่วงเวลา ซึ่งต่อมาได้มีการพัฒนาจากวิธีการของ Jackson โดยทำการประมาณขนาดของประชากรโดยอาศัยเทคนิคของการวิเคราะห์การถดถอยเพื่อสร้างสมการแสดงความสัมพันธ์ระหว่างขนาดของประชากรที่ประมาณขึ้นโดยใช้วิธีการของ Lincoln – Petersen ในการสุ่มแต่ละครั้ง กับตัวแปรเวลา แล้วจึงประมาณค่าขนาดของประชากรที่เวลาเริ่มต้นซึ่งได้ทำการสุ่มตัวอย่างในครั้งแรกและทำเครื่องหมายไว้ นั่นก็จะได้ว่าระยะตัดแกน Y เป็นค่าประมาณขนาดของประชากร

สมการถดถอยคือ

$$\hat{N}_i = b_0 + b_1 t_i \quad (5)$$

โดย $\hat{N}_i$ คือค่าประมาณขนาดของประชากรในการสุ่มครั้งที่ i , $i = 1, 2, \dots, n$

b_0 คือค่าประมาณของระยะตัดแกน Y (ค่าของ $\hat{N}$ เมื่อ $t_i = 0$)

b_1 คือค่าประมาณความชันของเส้นถดถอย

t_i คือเวลาที่ i , $i = 1, 2, \dots, n$

ในการวิจัยนี้ผู้วิจัยจะทำการพัฒนาวิธีการประมาณขนาดของประชากรขึ้นมาใหม่โดยอาศัยแนวคิดที่พัฒนามาจากวิธีของ Jackson(1939,1948) เนื่องจากพบว่าเป็นวิธีการที่สามารถดำเนินการได้สะดวกและไม่ยุ่งยากนั้นก็จะทำการสุ่มตัวอย่างในครั้งแรกและทำเครื่องหมายไว้ (M_1) หลังจากนั้นปล่อยกลับไปสู่ประชากร รอจนช่วงเวลาหนึ่งแล้วจึงทำการสุ่มตัวอย่างครั้งต่อ ๆ มา ณ ช่วงเวลาต่าง ๆ ที่เปลี่ยนแปลงไป โดยที่เมื่อสุ่มได้ในแต่ละครั้งแล้วให้นับจำนวนสมาชิกที่ถูกทำเครื่องหมายและสุ่มได้ในครั้งนั้น (m_i) แล้วจึงปล่อยคืนกลับไปสู่ประชากรอีกครั้งหนึ่ง ในกรณีนี้จะพบว่า $M_1 = M$ เพราะวิธีการนี้จะทำเครื่องหมายให้กับสมาชิกในการสุ่มตัวอย่างครั้งแรกเท่านั้น ซึ่งจากข้อสมมติที่ว่าประชากรเป็นประชากรแบบเปิด ดังนั้นจะพบว่าสัดส่วนของจำนวนสมาชิกที่ถูกทำเครื่องหมายและจับได้จะมีจำนวนลดลงเมื่อเวลาเปลี่ยนแปลงไปเพราะอาจมีสมาชิกที่ถูกทำเครื่องหมายตายหรืออพยพไปอยู่ในถิ่นอื่น ทำให้จำนวนสมาชิกที่ถูกทำเครื่องหมายและสุ่มได้ในครั้ง ต่อ ๆ มา (m_i) เมื่อเวลาผ่านไปมีจำนวนลดลง ดังนั้นถ้าทำการประมาณขนาดของประชากร

Lincoln – Petersen (1896,1930) ในการสุ่มแต่ละครั้ง กับตัวแปรเวลา แล้วจึงประมาณค่าขนาดของประชากรที่เวลาเริ่มต้นอาจทำให้ได้ค่าประมาณที่ต่ำกว่าความเป็นจริงมาก ผู้วิจัยจึงได้พัฒนาวิธีการประมาณขนาดของประชากรขึ้นมาใหม่โดยอาศัยแนวคิดของ Jackson ซึ่งในงานวิจัยครั้งนี้ได้พัฒนาขึ้นมาทั้งหมด 2 วิธีคือ

2.1 วิธีที่พัฒนาจาก Jackson โดยใช้ค่ามัธยฐานของเวลา (Modified Jackson using Median Time Method)

ในกรณีที่การสุ่มตัวอย่างดำเนินการในช่วงระยะเวลาที่ยาวนาน จะพบว่าเวลาจะเป็นตัวแปรสำคัญที่ใช้ในการประมาณค่าขนาดประชากร นั่นคือเวลาจะแปรผันตามขนาดของประชากรที่ประมาณขึ้น เนื่องจากจำนวนสมาชิกที่ถูกทำเครื่องหมายและสุ่มได้ในครั้งต่อ ๆ มา (m_i) จะมีแนวโน้มลดลง จากที่กล่าวข้างต้นจะพบว่าอาจเกิดปัญหาที่จะได้ค่าประมาณขนาดของประชากรที่ต่ำกว่าความเป็นจริงมากถ้าทำการประมาณ ณ ตำแหน่งของเวลาเริ่มต้น ดังนั้นผู้วิจัยจะเสนอแนวทางในการประมาณขนาดของประชากร ณ ตำแหน่งของเวลาที่เป็นตำแหน่งค่ากลางของช่วงเวลาที่ทำการสุ่มตัวอย่าง โดยอาศัยการหาค่ากลางโดยใช้ค่ามัธยฐาน (Median) เพื่อหาตำแหน่งของเวลาที่จะใช้ในการประมาณเพื่อให้ได้ค่าประมาณขนาดของประชากรที่ทำให้ความคลาดเคลื่อนต่ำที่สุด ดังนั้นวิธีการนี้จะสร้างสมการถดถอยเพื่อแสดงความสัมพันธ์ระหว่างขนาดของประชากรที่ประมาณขึ้นโดยใช้วิธีการของ Lincoln – Petersen ในการสุ่มแต่ละครั้ง กับตัวแปรเวลา แล้วจึงประมาณค่าขนาดของประชากรที่ตำแหน่งมัธยฐานของเวลาแทนที่เวลาเริ่มต้น

สมการถดถอยคือ

$$\hat{N}_i = b_0 + b_1 t_i \quad (6)$$

โดย $\hat{N}_i$ คือค่าประมาณขนาดของประชากรในการสุ่มครั้งที่ i , $i = 1, 2, \dots, n$

b_0 คือค่าประมาณของระยะตัดแกน Y

b_1 คือค่าประมาณความชันของเส้นถดถอย

t_i คือเวลาที่ i , $i = 1, 2, \dots, n$

2.2 วิธีที่พัฒนาจาก Jackson โดยใช้ค่ามัธยฐานของจำนวนของสมาชิกที่ถูกทำเครื่องหมายและสุ่มได้ในการสุ่มตัวอย่างครั้งถัดไป (Modified Jackson using Median m_i Method)

ในกรณีที่ช่วงเวลาที่ทำการสุ่มตัวอย่างมีระยะห่างไม่มากนัก อาจมีข้อสมมติว่าอัตราการเกิดหรือตาย และอัตราการอพยพย้ายถิ่นหรือการอพยพเข้ามาอยู่ในถิ่นใหม่มีสัดส่วนเท่า ๆ กันและเปลี่ยนแปลงไม่มากนัก ทำให้จำนวนสมาชิกที่ถูกทำเครื่องหมายไว้ใน การสุ่มตัวอย่างครั้งแรก (M) มีค่าไม่แตกต่างมากนักในการสุ่มตัวอย่างครั้งต่อ ๆ ไป ดังนั้นขนาดของประชากรที่ประมาณขึ้น อาจจะไม่ได้แปรผันตามเวลา ทำให้ตัวแปรเวลาอาจจะไม่ใช่ตัวแปรอิสระที่ดี แต่จะพบว่าขนาดของประชากรที่ประมาณได้จะมีความสัมพันธ์กับจำนวนสมาชิกที่ถูกทำเครื่องหมายไว้ในครั้งแรก และถูกสุ่มได้ในครั้งต่อ ๆ มา (m_i) ซึ่งอาจจะมีลำดับแบบสุ่ม (Random) เพราะไม่ได้ลดลงเมื่อเวลาเปลี่ยนแปลงไป ดังนั้นจึงอาจจะสร้างสมการถดถอยแสดงความสัมพันธ์ระหว่างขนาดของประชากรที่ประมาณได้ในการสุ่มตัวอย่างแต่ละครั้งกับจำนวนสมาชิกที่ถูกทำเครื่องหมายไว้ในครั้งแรกและถูกสุ่มในครั้งต่อ ๆ มา (m_i) ซึ่งจะมีความสัมพันธ์เชิงเส้นตรงและเป็นไปในลักษณะแปรผกผันกัน โดยมีข้อสมมติว่าเวลาที่ทำการศึกษาอยู่ในช่วงจำกัด และจะทำการประมาณค่าขนาดของประชากรที่ตำแหน่งมัธยฐานของ m_i สมการถดถอยคือ

$$\hat{N}_i = b_0 + b_1 m_i \quad (7)$$

โดย $\hat{N}_i$ คือค่าประมาณขนาดของประชากรในการสุ่มครั้งที่ i , $i = 1, 2, \dots, n$

b_0 คือค่าประมาณของระยะตัดแกน Y

b_1 คือค่าประมาณความชันของเส้นถดถอย

m_i คือจำนวนของสมาชิกที่ถูกทำเครื่องหมาย และสุ่มได้ในการสุ่มตัวอย่างครั้งที่ i ,

$i = 1, 2, \dots, n$

บทที่ 3

วิธีดำเนินงานวิจัย

ในการวิจัยเรื่องการประมาณขนาดของประชากรโดยวิธี Capture – Recapture ในกรณีที่เป็นการสุ่มตัวอย่างแบบหลายครั้ง ได้ทำการนำเสนอวิธีการประมาณขนาดของประชากรโดยวิธี Capture – Recapture กรณีที่เป็นการสุ่มตัวอย่างแบบหลายครั้งโดยอาศัยแนวคิดที่พัฒนามาจากวิธีของ Jackson ขึ้นใหม่ทั้งหมด 2 วิธีคือ วิธีที่พัฒนามาจาก Jackson โดยใช้ค่ามัธยฐานของเวลา (Modified Jackson using Median Time Method) และวิธีที่พัฒนามาจาก Jackson โดยใช้ค่ามัธยฐานของจำนวนของสมาชิกที่ถูกทำเครื่องหมายและสุ่มได้ในการสุ่มตัวอย่างครั้งถัดไป (Modified Jackson using Median m_1 Method) และจะทำการเปรียบเทียบวิธีการประมาณขนาดของประชากรดังกล่าวข้างต้นกับวิธีเดิมคือวิธีที่พัฒมาจากเจ็คสัน (Modified Jackson Method) รวมทั้งทำการประมาณค่าแบบช่วงของจำนวนของสมาชิกที่ถูกทำเครื่องหมายและสุ่มได้ในการสุ่มตัวอย่างครั้งถัดไปที่เหมาะสม โดยจะศึกษาการประมาณขนาดของประชากรโดยวิธี Capture – Recapture สำหรับประชากรแบบเปิด เฉพาะกรณีที่เป็นการสุ่มตัวอย่างแบบหลายครั้ง ซึ่งทำเครื่องหมายให้กับสมาชิกแต่ละตัวที่สุ่มได้ในการสุ่มตัวอย่างครั้งแรกเท่านั้น การศึกษานี้จะกำหนดค่า $N = 500$ และ $1,000$ โดยแบ่งออกเป็นกรณีที่ M และ n มีค่าเท่ากันและมีค่าแตกต่างกัน โดยในแต่ละกรณีทำซ้ำ $10,000$ ครั้ง ซึ่งผู้วิจัยจะนำผลการจำลองของทั้ง 3 วิธีดังกล่าวมาทำการเปรียบเทียบโดยพิจารณาจากค่าเฉลี่ยของเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์เป็นเกณฑ์ว่าวิธีการประมาณขนาดของประชากรทั้ง 3 วิธีนี้ วิธีใดที่ให้ค่าเฉลี่ยของเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์ต่ำสุดจะเลือกวิธีการนั้นเป็นวิธีการประมาณค่าต่อไป โดยมีขั้นตอนในการดำเนินงานวิจัยดังต่อไปนี้

ขั้นที่ 1 ศึกษาเอกสารและงานวิจัยที่เกี่ยวข้อง

ขั้นที่ 2 ศึกษาทฤษฎีที่เกี่ยวข้องกับการประมาณขนาดของประชากรโดยวิธี Capture – Recapture สำหรับประชากรแบบเปิด กรณีที่เป็นการสุ่มตัวอย่างแบบหลายครั้ง

ขั้นที่ 3 ศึกษาวิธีการวิเคราะห์การถดถอยเพื่อนำมาประยุกต์ใช้ในการประมาณขนาดของประชากรโดยวิธี Capture – Recapture

ขั้นที่ 4 เสนอวิธีการประมาณขนาดของประชากรขึ้นใหม่โดยอาศัยแนวคิดที่พัฒนามาจากวิธีของ Jackson ซึ่งในงานวิจัยครั้งนี้ได้พัฒนาขึ้นมาทั้งหมด 2 วิธีคือ

1. วิธีที่พัฒนามาจาก Jackson โดยใช้ค่ามัธยฐานของเวลา (Modified Jackson using Median Time Method)
2. วิธีที่พัฒนามาจาก Jackson โดยใช้ค่ามัธยฐานของจำนวนของสมาชิกที่ถูกทำเครื่องหมายและสุ่มได้ในการสุ่มตัวอย่างครั้งถัดไป (Modified Jackson using Median m_i Method)

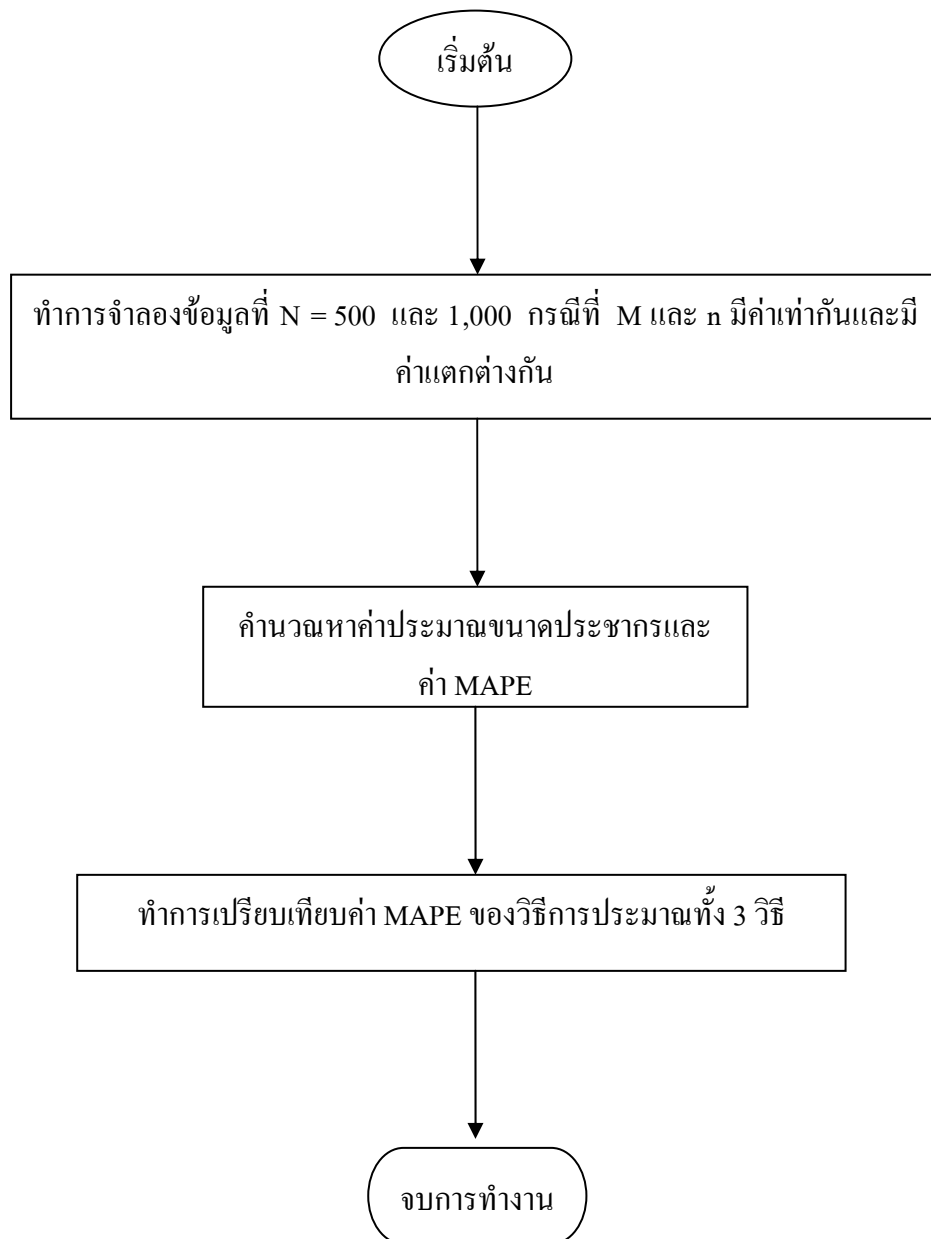
ขั้นที่ 5 เขียนโปรแกรม

ในการวิจัยครั้งนี้ได้ทำการจำลองค่าประชากร $N = 500$ และ $1,000$ โดยแบ่งออกเป็นกรณีที่ M และ n มีค่าเท่ากันและมีค่าแตกต่างกัน โดยในแต่ละกรณีทำซ้ำ $10,000$ ครั้ง โดยใช้โปรแกรม Minitab Macros

ขั้นที่ 6 ทดสอบโปรแกรมและทำการแก้ไข

ขั้นที่ 7 ปรับปรุงโปรแกรม

ขั้นที่ 8 ประมวลผลการวิจัย



รูปที่ 3.1 ขั้นตอนการจำลองข้อมูล

บทที่ 4

ผลการวิเคราะห์ข้อมูล

จากการศึกษาวิจัยในครั้งนี้ ได้ดำเนินการตามขั้นตอนการดำเนินงานวิจัยเพื่อทำการประมาณขนาดของประชากรโดยวิธี Capture – Recapture ในกรณีที่ประชากรแบบเปิดและเป็นการสุ่มตัวอย่างแบบหลายครั้งขึ้นใหม่ เพื่อให้ได้ค่าประมาณขนาดของประชากรที่มีความคลาดเคลื่อนต่ำที่สุด โดยในการศึกษานี้จะกำหนดค่า $N = 500$ และ $1,000$ โดยแบ่งออกเป็นกรณีที่ M และ n มีค่าเท่ากันและมีค่าแตกต่างกัน โดยในแต่ละกรณีทำซ้ำ $10,000$ ครั้ง ซึ่งผู้วิจัยจะนำผลการจำลองของทั้ง 3 วิธีดังกล่าวมาทำการเปรียบเทียบ โดยพิจารณาจากค่าเฉลี่ยของเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์เป็นเกณฑ์ว่าวิธีการประมาณขนาดของประชากรทั้ง 3 วิธีนี้ วิธีใดที่ให้ค่าเฉลี่ยของเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์ต่ำสุด (MAPE) จะเลือกวิธีการนั้นเป็นวิธีการประมาณค่าต่อไป ดังจะเสนอผลการวิเคราะห์ข้อมูลดังต่อไปนี้

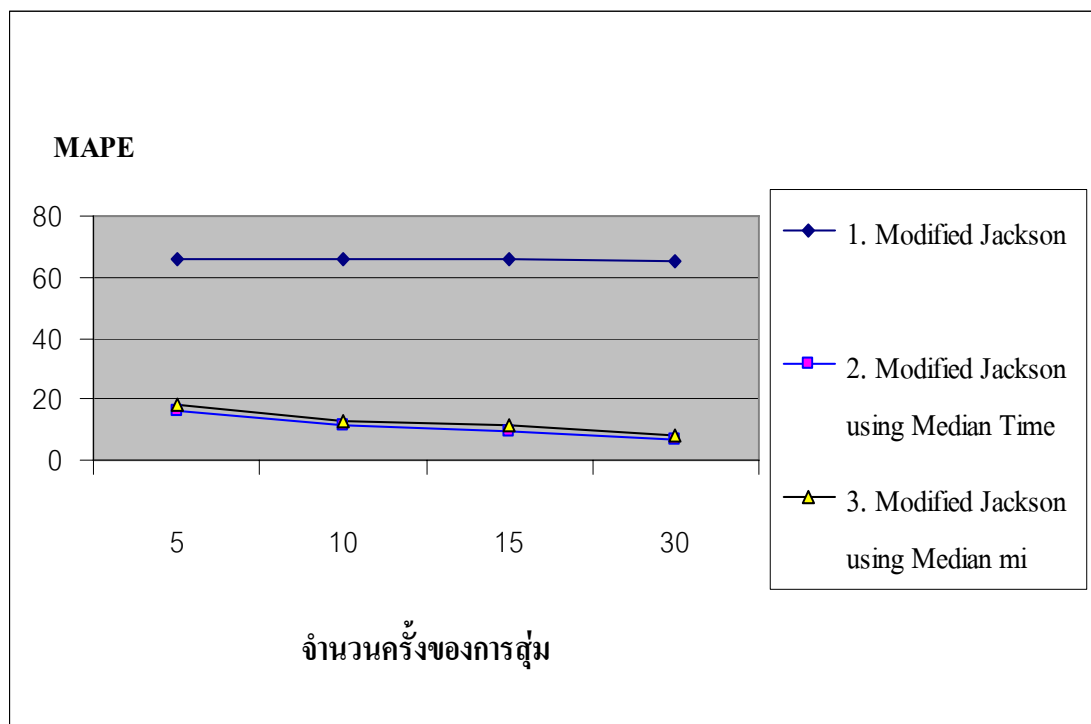
ตารางที่ 4.1 ค่า MAPE สำหรับจำนวนครั้งของการสุ่มที่แตกต่างกัน กรณีประชากรขนาด $N = 500$ และขนาดตัวอย่างเหมือนกัน

M	n	f1 = M/N	f2 = n/N	วิธีการ	MAPE สำหรับจำนวนครั้งของการสุ่มที่แตกต่างกัน			
					5	10	15	30
50	50	0.1	0.1	1. Modified Jackson	65.86	65.57	65.58	65.48
				2. Modified Jackson using Median Time	15.86	11.66	9.52	6.94
				3. Modified Jackson using Median m_i	18.12	13.08	11.74	8.21
100	100	0.2	0.2	1. Modified Jackson	29.99	29.85	29.99	29.95
				2. Modified Jackson using Median Time	6.64	4.64	3.81	2.68
				3. Modified Jackson using Median m_i	7.78	5.37	4.70	3.23
200	200	0.4	0.4	1. Modified Jackson	11.33	11.33	11.29	11.35
				2. Modified Jackson using Median Time	2.43	1.69	1.39	0.99
				3. Modified Jackson using Median m_i	2.92	1.99	1.72	1.20
400	400	0.8	0.8	1. Modified Jackson	1.89	1.89	1.89	1.89
				2. Modified Jackson using Median Time	0.39	0.28	0.23	0.16
				3. Modified Jackson using Median m_i	0.48	0.34	0.28	0.20

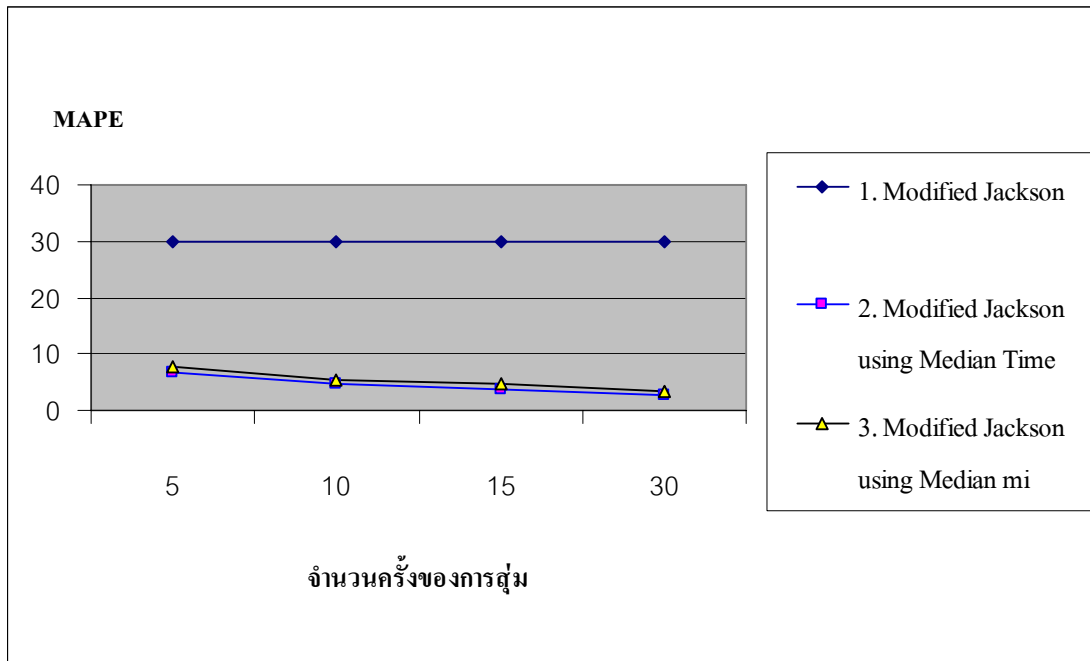
จากตารางที่ 4.1 เมื่อพิจารณาค่า MAPE สำหรับจำนวนครั้งของการสุ่มที่แตกต่างกัน ในกรณีประชากรขนาด $N = 500$ และขนาดตัวอย่างเหมือนกัน พบว่าวิธีที่พัฒนาจาก Jackson โดยใช้ค่ามัธยฐานของเวลา (Modified Jackson using Median Time) จะให้ค่า MAPE ต่ำที่สุด รองลงมาคือวิธีที่พัฒนาจาก Jackson โดยใช้ค่ามัธยฐานของจำนวนของสมาชิกที่ถูกทำเครื่องหมายและสุ่มได้ในการสุ่มตัวอย่างครั้งถัดไป (Modified Jackson using Median m_i) และวิธีที่พัฒนาจาก Jackson (Modified Jackson) ตามลำดับ โดยจะพบว่าถ้าสัดส่วนการสุ่มตัวอย่างและจำนวนครั้งของการสุ่มตัวอย่างเพิ่มขึ้น ค่า MAPE จะมีค่าต่ำลง ยกเว้นวิธีที่พัฒนาจาก Jackson จะพบว่าค่า MAPE สำหรับจำนวนครั้งของการสุ่มที่แตกต่างกันจะมีค่าใกล้เคียงกัน ดังนั้นวิธีการที่เหมาะสมในการประมาณขนาดของประชากรโดยวิธี Capture – Recapture กรณีที่เป็นการสุ่มตัวอย่างแบบหลายครั้งคือวิธีที่พัฒนาจาก Jackson โดยใช้ค่ามัธยฐานของเวลา

จากข้อมูลในตารางที่ 4.1 นำเสนอรูปแสดงการเปรียบเทียบค่าเฉลี่ยของเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์ (MAPE) ของวิธีการต่าง ๆ สำหรับจำนวนครั้งของการสุ่มที่แตกต่างกัน กรณีประชากรขนาด $N = 500$ และขนาดตัวอย่างเหมือนกัน ดังแสดงในรูปที่ 4.1 ถึง 4.4 ตามลำดับ

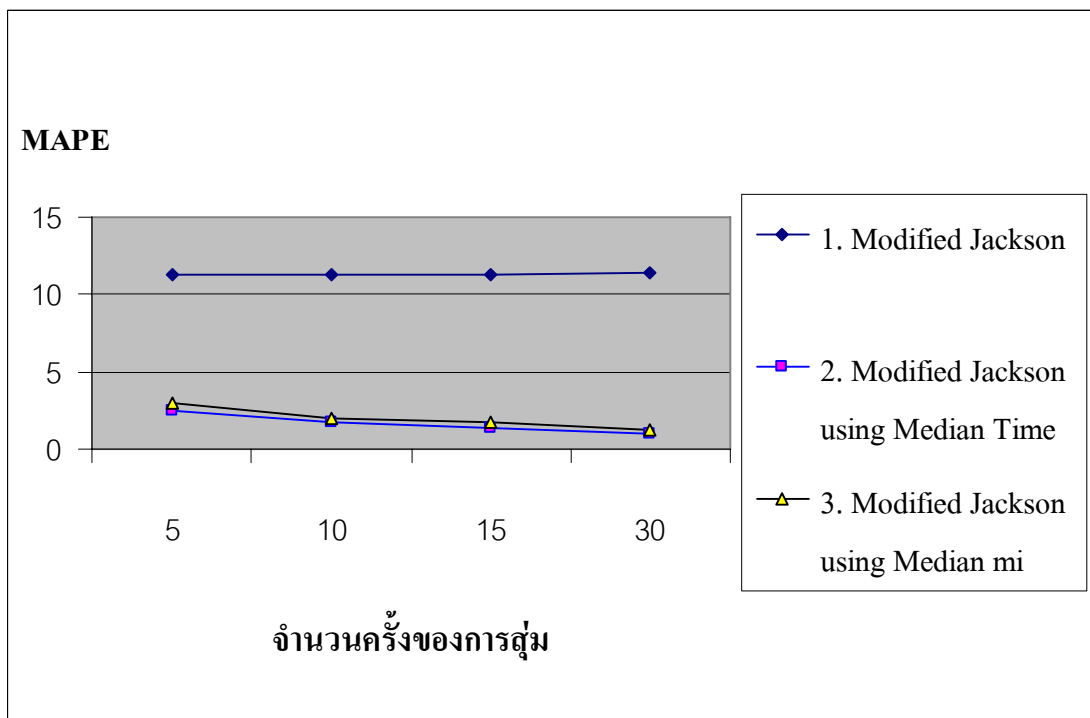
รูปที่ 4.1 ค่าเฉลี่ยของเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์ (MAPE) ของวิธีการต่าง ๆ สำหรับจำนวนครั้งของการสุ่มที่แตกต่างกัน กรณีประชากรขนาด $N = 500$ และ $M = n = 50$



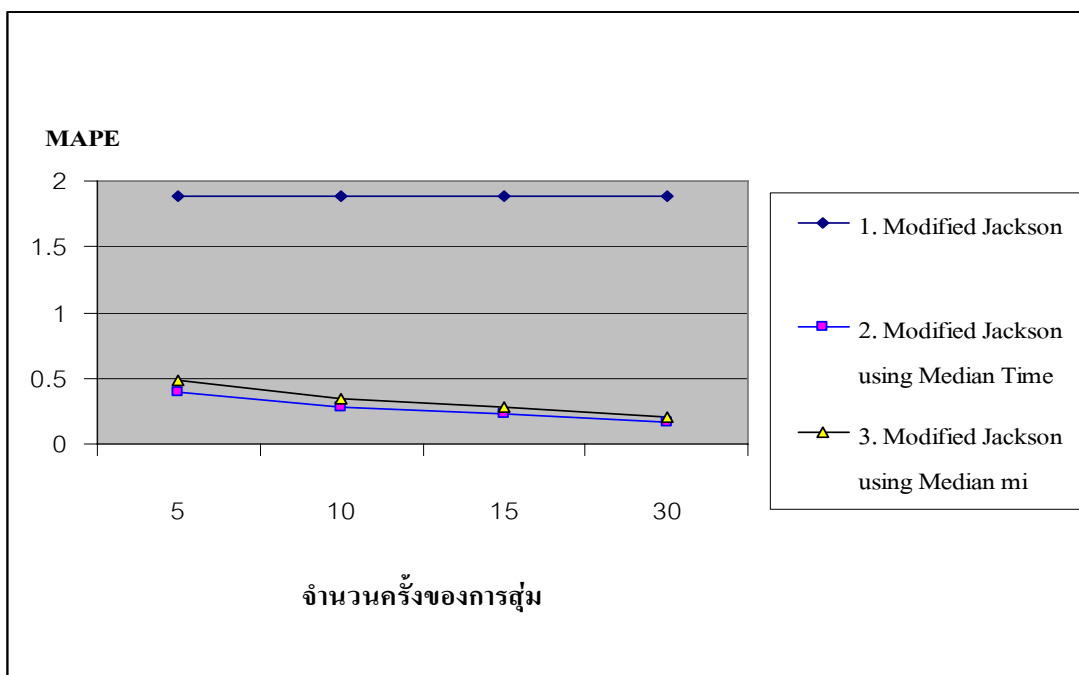
รูปที่ 4.2 ค่าเฉลี่ยของเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์ (MAPE) ของวิธีการต่าง ๆ สำหรับจำนวนครั้งของการสุ่มที่แตกต่างกัน กรณีประชากรขนาด $N = 500$ และ $M = n = 100$



รูปที่ 4.3 ค่าเฉลี่ยของเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์ (MAPE) ของวิธีการต่าง ๆ สำหรับจำนวนครั้งของการสุ่มที่แตกต่างกัน กรณีประชากรขนาด $N = 500$ และ $M = n = 200$



รูปที่ 4.4 ค่าเฉลี่ยของเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์ (MAPE) ของวิธีการต่าง ๆ สำหรับจำนวนครั้งของการสุ่มที่แตกต่างกัน กรณีประชากรขนาด $N = 500$ และ $M = n = 400$



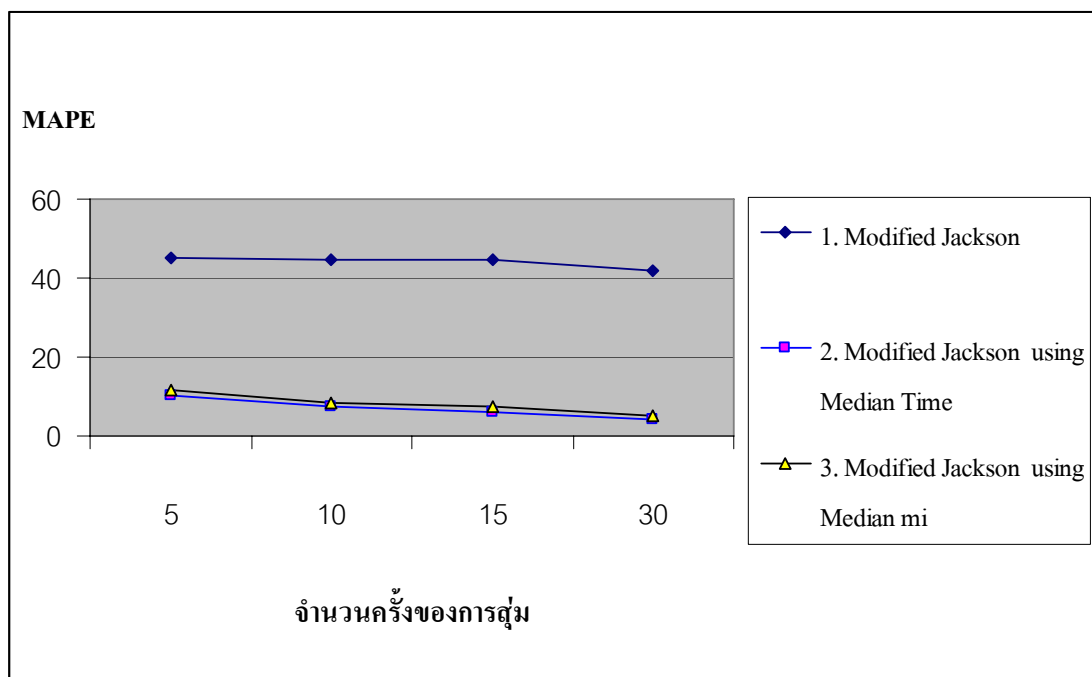
ตารางที่ 4.2 ค่า MAPE สำหรับจำนวนครั้งของการสุ่มที่แตกต่างกัน กรณีประชากรขนาด $N = 500$ และขนาดตัวอย่างแตกต่างกัน

M	n	f1 = M/N	f2 = n/N	Methods	MAPE สำหรับจำนวนครั้งของการสุ่มที่แตกต่างกัน			
					5	10	15	30
50	100	0.1	0.2	1. Modified Jackson	44.97	44.81	44.65	42.08
				2. Modified Jackson using Median Time	10.19	7.23	5.93	4.22
				3. Modified Jackson using Median m_i	11.81	8.33	7.31	5.15
100	50	0.2	0.1	1. Modified Jackson	44.41	44.72	44.59	44.57
				2. Modified Jackson using Median Time	10.14	7.23	5.92	4.25
				3. Modified Jackson using Median m_i	11.85	8.24	7.29	5.11
100	200	0.2	0.4	1. Modified Jackson	18.43	18.49	18.50	18.45
				2. Modified Jackson using Median Time	3.95	2.76	2.30	1.62
				3. Modified Jackson using Median m_i	4.67	3.25	2.85	1.98
200	100	0.4	0.2	1. Modified Jackson	18.39	18.40	18.45	18.47
				2. Modified Jackson using Median Time	3.98	2.81	2.27	1.61
				3. Modified Jackson using Median m_i	4.71	3.29	2.82	1.97

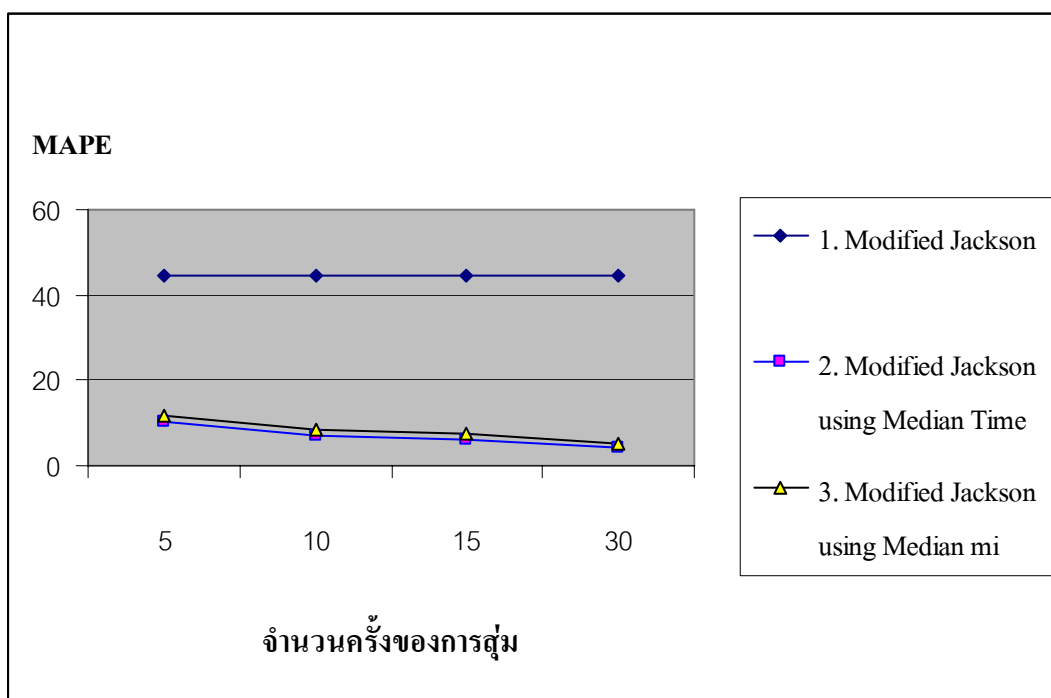
จากตารางที่ 4.2 เมื่อพิจารณาค่า MAPE สำหรับจำนวนครั้งของการสุ่มที่แตกต่างกัน ในกรณีประชากรขนาด $N = 500$ และขนาดตัวอย่างแตกต่างกัน พบว่าวิธีที่พัฒนาจาก Jackson โดยใช้ค่ามัธยฐานของเวลา (Modified Jackson using Median Time) จะให้ค่า MAPE ต่ำที่สุด รองลงมาคือวิธีที่พัฒนาจาก Jackson โดยใช้ค่ามัธยฐานของจำนวนของสมาชิกที่ถูกทำเครื่องหมายและสุ่มได้ในการสุ่มตัวอย่างครั้งถัดไป (Modified Jackson using Median m_i) และวิธีที่พัฒนาจาก Jackson (Modified Jackson) ตามลำดับ โดยจะพบว่าถ้าสัดส่วนการสุ่มตัวอย่างและจำนวนครั้งของการสุ่มตัวอย่างเพิ่มขึ้น ค่า MAPE จะมีค่าต่ำลง ยกเว้นวิธีที่พัฒนาจาก Jackson จะพบว่าค่า MAPE สำหรับจำนวนครั้งของการสุ่มที่แตกต่างกันจะมีค่าใกล้เคียงกัน ดังนั้นวิธีการที่เหมาะสมในการประมาณขนาดของประชากรโดยวิธี Capture – Recapture กรณีที่เป็นการสุ่มตัวอย่างแบบหลายครั้งคือวิธีที่พัฒนาจาก Jackson โดยใช้ค่ามัธยฐานของเวลา

จากข้อมูลในตารางที่ 4.2 นำเสนอรูปแสดงการเปรียบเทียบค่าเฉลี่ยของเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์ (MAPE) ของวิธีการต่าง ๆ สำหรับจำนวนครั้งของการสุ่มที่แตกต่างกัน กรณีประชากรขนาด $N = 500$ และขนาดตัวอย่างแตกต่างกัน ดังแสดงในรูปที่ 4.5 ถึง 4.8 ตามลำดับ

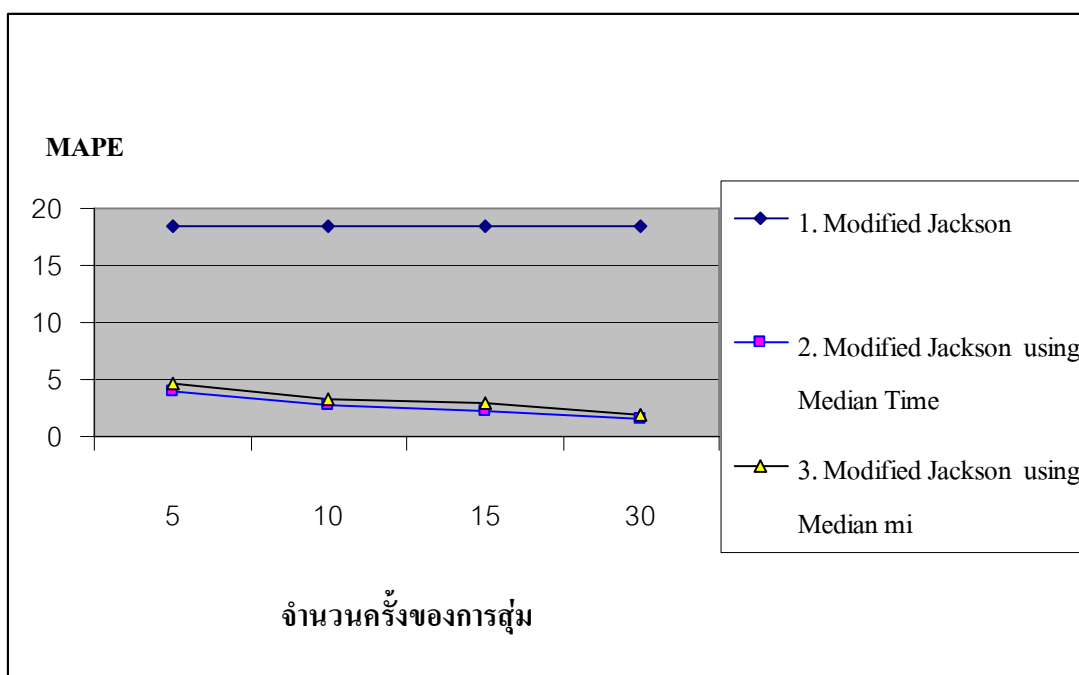
รูปที่ 4.5 ค่าเฉลี่ยของเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์ (MAPE) ของวิธีการต่าง ๆ สำหรับจำนวนครั้งของการสุ่มที่แตกต่างกัน กรณีประชากรขนาด $N = 500$ และ $M = 50$, $n = 100$



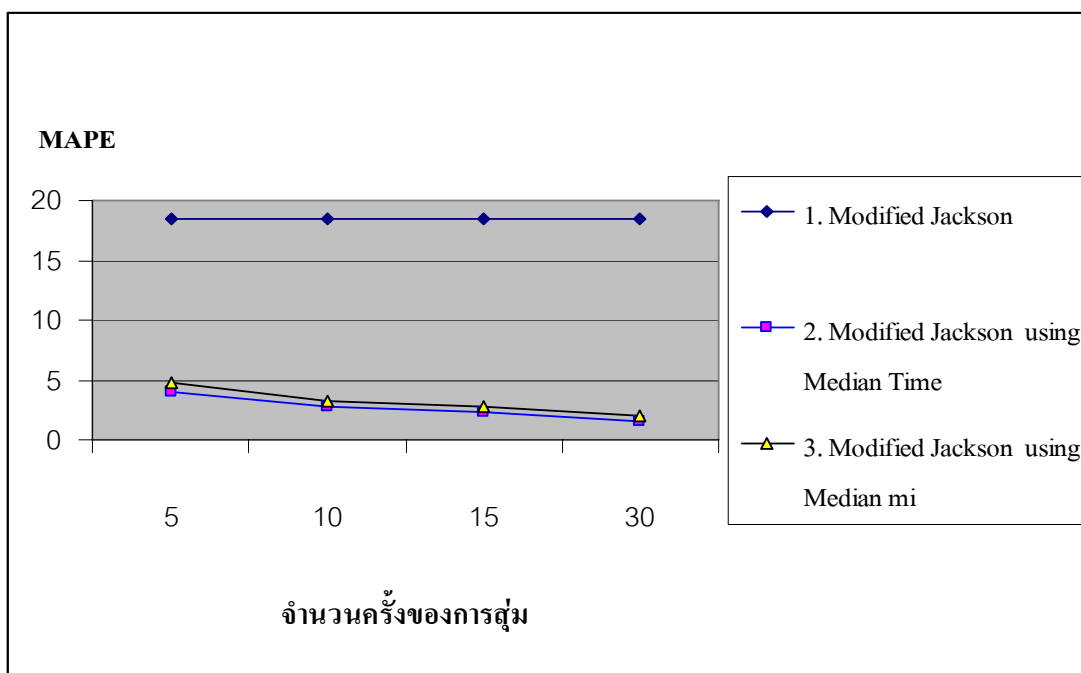
รูปที่ 4.6 ค่าเฉลี่ยของเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์ (MAPE) ของวิธีการต่าง ๆ สำหรับ
จำนวนครั้งของการสุ่มที่แตกต่างกัน กรณีประชากรขนาด $N = 500$ และ $M = 100$, $n = 50$



รูปที่ 4.7 ค่าเฉลี่ยของเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์ (MAPE) ของวิธีการต่าง ๆ สำหรับ
จำนวนครั้งของการสุ่มที่แตกต่างกัน กรณีประชากรขนาด $N = 500$ และ $M = 100$, $n = 200$



รูปที่ 4.8 ค่าเฉลี่ยของเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์ (MAPE) ของวิธีการต่าง ๆ สำหรับจำนวนครั้งของการสุ่มที่แตกต่างกัน กรณีประชากรขนาด $N = 500$ และ $M = 200$, $n = 100$



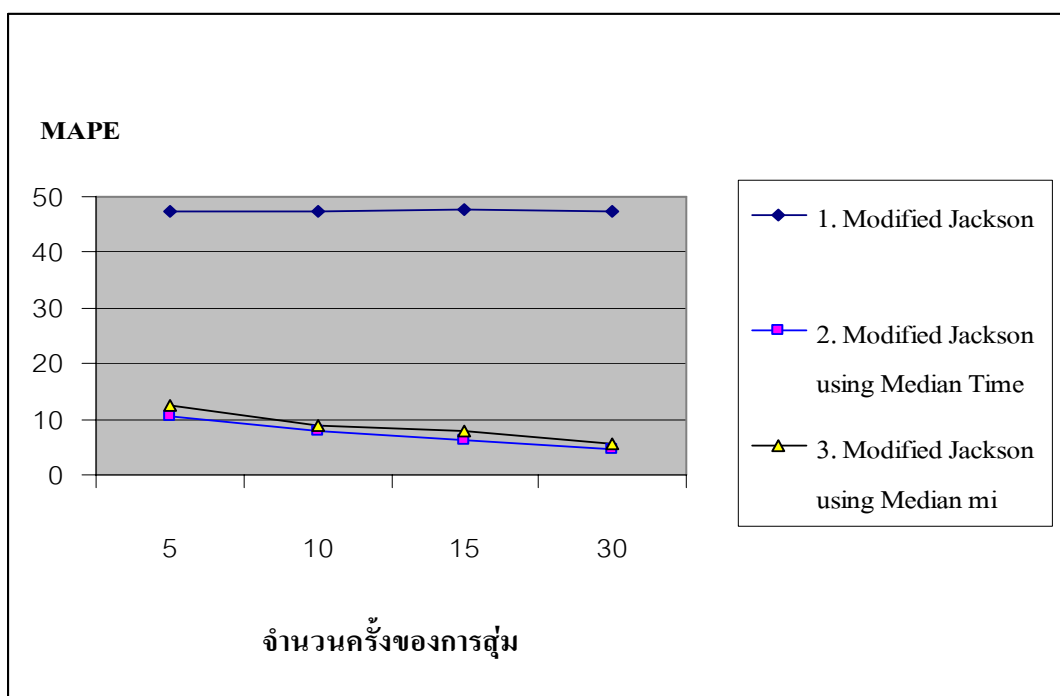
ตารางที่ 4.3 ค่า MAPE สำหรับจำนวนครั้งของการสุ่มที่แตกต่างกัน กรณีประชากรขนาด $N = 1,000$ และขนาดตัวอย่างเหมือนกัน

M	n	f1 = M/N	f2 =n/N	Methods	MAPE สำหรับจำนวนครั้งของการสุ่มที่แตกต่างกัน			
					5	10	15	30
100	100	0.1	0.1	1. Modified Jackson	47.50	47.50	47.56	47.46
				2. Modified Jackson using Median Time	10.66	7.76	6.35	4.54
				3. Modified Jackson using Median m_i	12.47	8.82	7.80	5.49
200	200	0.2	0.2	1. Modified Jackson	21.16	21.25	21.29	21.24
				2. Modified Jackson using Median Time	4.61	3.21	2.62	1.88
				3. Modified Jackson using Median m_i	5.51	3.74	3.25	2.28
400	400	0.4	0.4	1. Modified Jackson	8.05	8.01	8.00	8.01
				2. Modified Jackson using Median Time	1.71	1.22	0.97	0.70
				3. Modified Jackson using Median m_i	2.02	1.42	1.21	0.85
800	800	0.8	0.8	1. Modified Jackson	1.33	1.33	1.34	1.33
				2. Modified Jackson using Median Time	0.28	0.20	0.16	0.12
				3. Modified Jackson using Median m_i	0.34	0.24	0.20	0.14

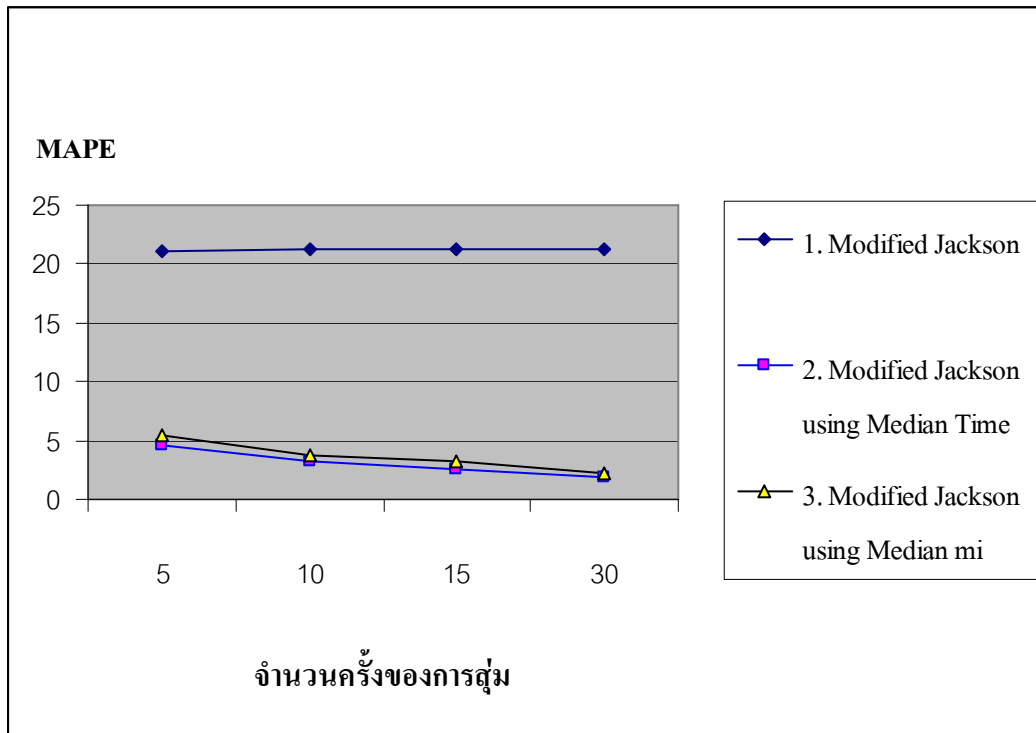
จากตารางที่ 4.3 เมื่อพิจารณาค่า MAPE สำหรับจำนวนครั้งของการสุ่มที่แตกต่างกัน ในกรณีประชากรขนาด $N = 1,000$ และขนาดตัวอย่างเหมือนกัน พบว่าวิธีที่พัฒนาจาก Jackson โดยใช้ค่ามัธยฐานของเวลา (Modified Jackson using Median Time) จะให้ค่า MAPE ต่ำที่สุด รองลงมาคือวิธีที่พัฒนาจาก Jackson โดยใช้ค่ามัธยฐานของจำนวนของสมาชิกที่ถูกทำเครื่องหมายและสุ่มได้ในการสุ่มตัวอย่างครั้งถัดไป (Modified Jackson using Median m_i) และวิธีที่พัฒนาจาก Jackson (Modified Jackson) ตามลำดับ โดยจะพบว่าถ้าสัดส่วนการสุ่มตัวอย่างและจำนวนครั้งของการสุ่มตัวอย่างเพิ่มขึ้น ค่า MAPE จะมีค่าต่ำลง ยกเว้นวิธีที่พัฒนาจาก Jackson จะพบว่าค่า MAPE สำหรับจำนวนครั้งของการสุ่มที่แตกต่างกันจะมีค่าใกล้เคียงกัน ดังนั้นวิธีการที่เหมาะสมในการประมาณขนาดของประชากรโดยวิธี Capture – Recapture กรณีที่เป็นการสุ่มตัวอย่างแบบหลายครั้งคือวิธีที่พัฒนาจาก Jackson โดยใช้ค่ามัธยฐานของเวลา

จากข้อมูลในตารางที่ 4.3 นำเสนอรูปแสดงการเปรียบเทียบค่าเฉลี่ยของเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์ (MAPE) ของวิธีการต่าง ๆ สำหรับจำนวนครั้งของการสุ่มที่แตกต่างกัน กรณีประชากรขนาด $N = 1,000$ และขนาดตัวอย่างเหมือนกัน ดังแสดงในรูปที่ 4.9 ถึง 4.12 ตามลำดับ

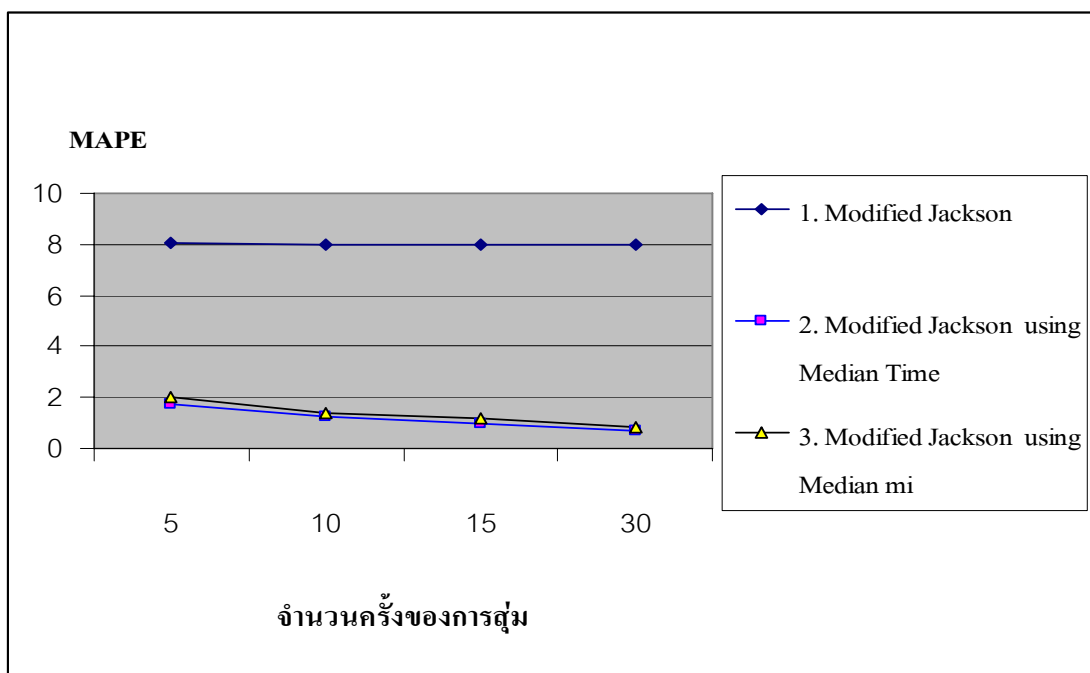
รูปที่ 4.9 ค่าเฉลี่ยของเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์ (MAPE) ของวิธีการต่าง ๆ สำหรับจำนวนครั้งของการสุ่มที่แตกต่างกัน กรณีประชากรขนาด $N = 1,000$ และ $M = n = 100$



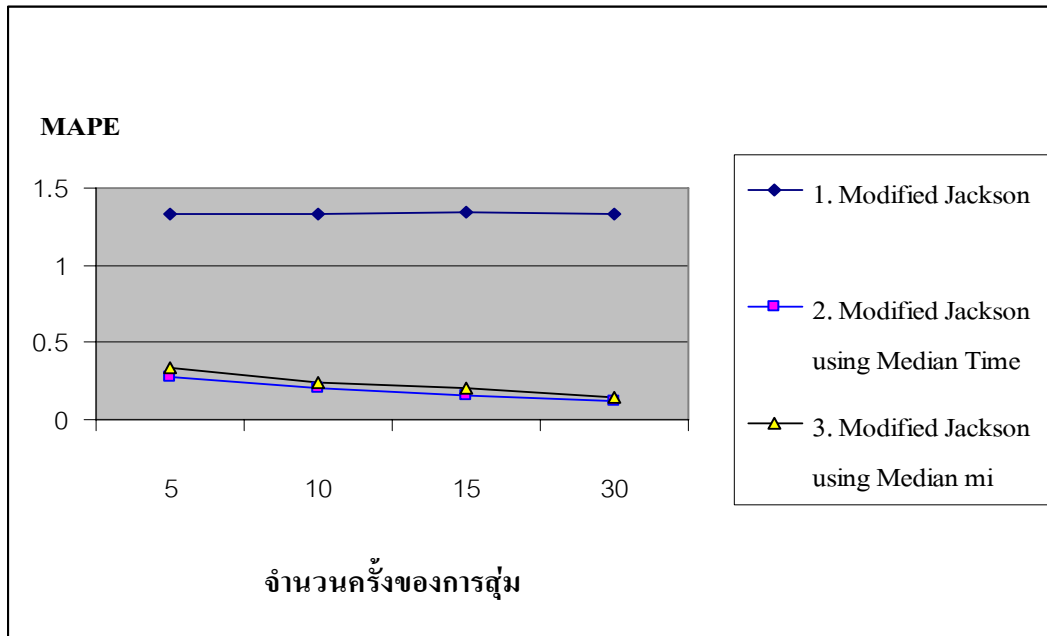
รูปที่ 4.10 ค่าเฉลี่ยของเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์ (MAPE) ของวิธีการต่าง ๆ สำหรับจำนวนครั้งของการสุ่มที่แตกต่างกัน กรณีประชากรขนาด $N = 1,000$ และ $M = n = 200$



รูปที่ 4.11 ค่าเฉลี่ยของเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์ (MAPE) ของวิธีการต่าง ๆ สำหรับจำนวนครั้งของการสุ่มที่แตกต่างกัน กรณีประชากรขนาด $N = 500$ และ $M = n = 400$



รูปที่ 4.12 ค่าเฉลี่ยของเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์ (MAPE) ของวิธีการต่าง ๆ สำหรับจำนวนครั้งของการสุ่มที่แตกต่างกัน กรณีประชากรขนาด $N = 500$ และ $M = n = 800$



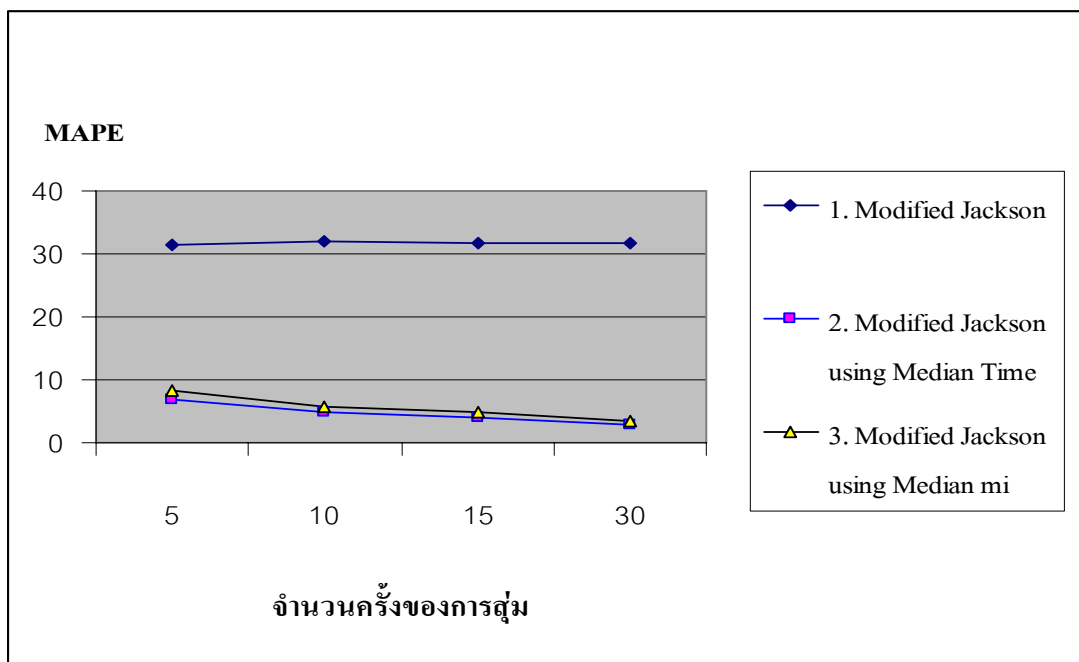
ตารางที่ 4.4 ค่า MAPE สำหรับจำนวนครั้งของการสุ่มที่แตกต่างกัน กรณีประชากรขนาด $N = 1,000$ และขนาดตัวอย่างแตกต่างกัน

M	n	f1 = M/N	f2 = n/N	Methods	MAPE สำหรับจำนวนครั้งของการสุ่มที่แตกต่างกัน			
					5	10	15	30
100	200	0.1	0.2	1. Modified Jackson	31.57	31.86	31.68	31.72
				2. Modified Jackson using Median Time	6.90	4.98	4.03	2.87
				3. Modified Jackson using Median m_i	8.16	5.74	4.97	3.50
200	100	0.2	0.1	1. Modified Jackson	31.90	31.68	31.77	31.82
				2. Modified Jackson using Median Time	6.96	4.89	4.10	2.85
				3. Modified Jackson using Median m_i	8.17	5.70	4.97	3.49
200	400	0.2	0.4	1. Modified Jackson	13.08	13.05	13.04	13.05
				2. Modified Jackson using Median Time	2.79	1.96	1.60	1.15
				3. Modified Jackson using Median m_i	3.30	2.30	1.99	1.40
400	200	0.4	0.2	1. Modified Jackson	13.11	13.03	13.07	13.07
				2. Modified Jackson using Median Time	2.77	1.96	1.61	1.13
				3. Modified Jackson using Median m_i	3.34	2.31	2.00	1.38

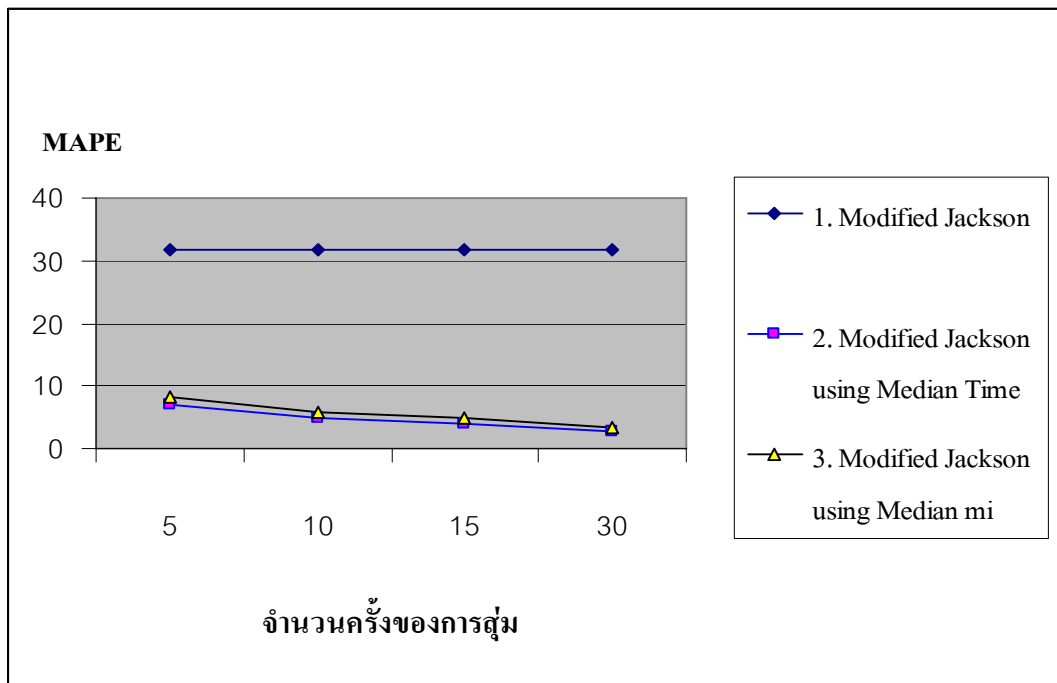
จากตารางที่ 4.4 เมื่อพิจารณาค่า MAPE สำหรับจำนวนครั้งของการสุ่มที่แตกต่างกัน ในกรณีประชากรขนาด $N = 1,000$ และขนาดตัวอย่างแตกต่างกัน พบว่าวิธีที่พัฒนาจาก Jackson โดยใช้ค่ามัธยฐานของเวลา (Modified Jackson using Median Time) จะให้ค่า MAPE ต่ำที่สุด รองลงมาคือวิธีที่พัฒนาจาก Jackson โดยใช้ค่ามัธยฐานของจำนวนของสมาชิกที่ถูกทำเครื่องหมาย และสุ่มได้ในการสุ่มตัวอย่างครั้งถัดไป (Modified Jackson using Median m_i) และวิธีที่พัฒนาจาก Jackson (Modified Jackson) ตามลำดับ โดยจะพบว่าถ้าสัดส่วนการสุ่มตัวอย่างและจำนวนครั้งของการสุ่มตัวอย่างเพิ่มขึ้น ค่า MAPE จะมีค่าต่ำลง ยกเว้นวิธีที่พัฒนาจาก Jackson จะพบว่าค่า MAPE สำหรับจำนวนครั้งของการสุ่มที่แตกต่างกันจะมีค่าใกล้เคียงกัน ดังนั้นวิธีการที่เหมาะสมในการประมาณขนาดของประชากรโดยวิธี Capture – Recapture กรณีที่เป็นการสุ่มตัวอย่างแบบหลายครั้งคือวิธีที่พัฒนาจาก Jackson โดยใช้ค่ามัธยฐานของเวลา

จากข้อมูลในตารางที่ 4.4 นำเสนอรูปแสดงการเปรียบเทียบค่าเฉลี่ยของเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์ (MAPE) ของวิธีการต่าง ๆ สำหรับจำนวนครั้งของการสุ่มที่แตกต่างกัน กรณีประชากรขนาด $N = 1,000$ และขนาดตัวอย่างแตกต่างกัน ดังแสดงในรูปที่ 4.13 ถึง 4.16 ตามลำดับ

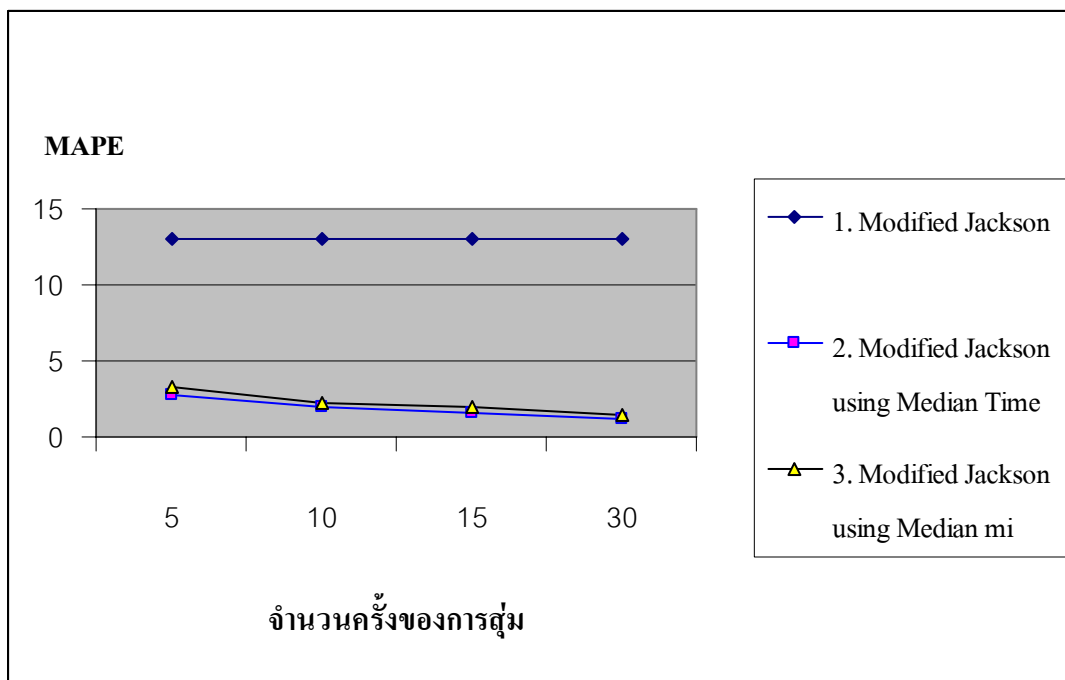
รูปที่ 4.13 ค่าเฉลี่ยของเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์ (MAPE) ของวิธีการต่าง ๆ สำหรับจำนวนครั้งของการสุ่มที่แตกต่างกัน กรณีประชากรขนาด $N = 1,000$ และ $M = 100$, $n = 200$



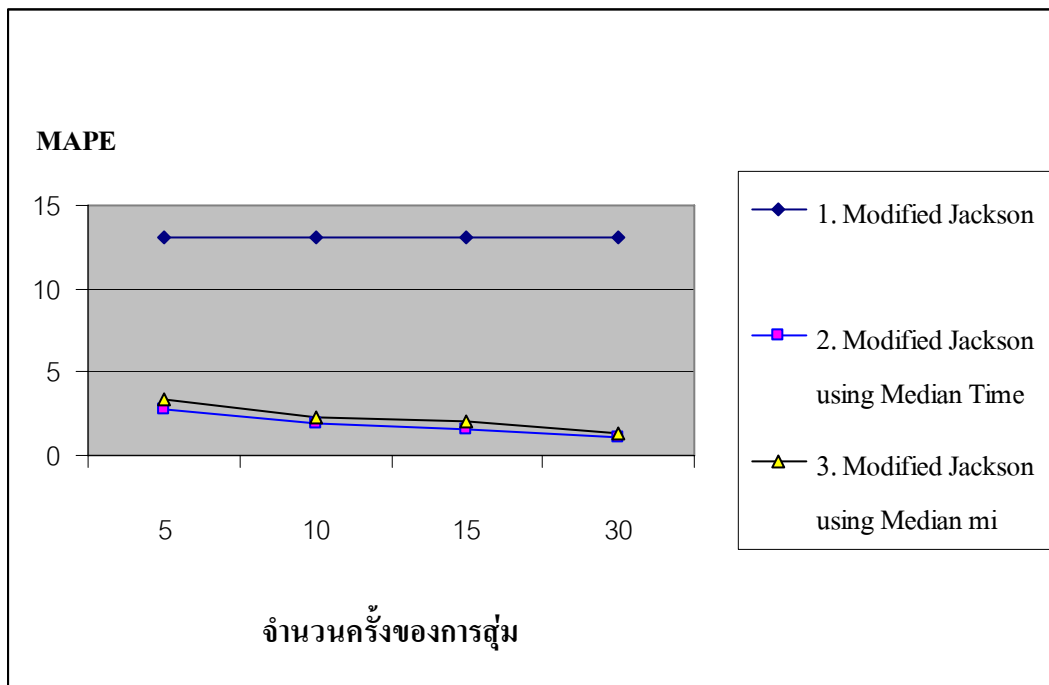
รูปที่ 4.14 ค่าเฉลี่ยของเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์ (MAPE) ของวิธีการต่าง ๆ สำหรับจำนวนครั้งของการสุ่มที่แตกต่างกัน กรณีประชากรขนาด $N = 1,000$ และ $M = 200$, $n = 100$



รูปที่ 4.15 ค่าเฉลี่ยของเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์ (MAPE) ของวิธีการต่าง ๆ สำหรับจำนวนครั้งของการสุ่มที่แตกต่างกัน กรณีประชากรขนาด $N = 1,000$ และ $M = 200$, $n = 400$



รูปที่ 4.16 ค่าเฉลี่ยของเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์ (MAPE) ของวิธีการต่าง ๆ สำหรับจำนวนครั้งของการสุ่มที่แตกต่างกัน กรณีประชากรขนาด $N = 1,000$ และ $M = 400$, $n = 200$



ตารางที่ 4.5 ช่วงความเชื่อมั่น 95% ของจำนวนของสมาชิกที่ถูกทำเครื่องหมายและสุ่มได้ในการสุ่มตัวอย่างครั้งถัดไป กรณีประชากรขนาด $N = 500$ และขนาดตัวอย่างเหมือนกัน

M	n	f1 = M/N	f2 = n/N	ช่วงความเชื่อมั่น 95% ของจำนวนของสมาชิกที่ถูกทำเครื่องหมายและสุ่มได้ในการสุ่มตัวอย่างครั้งถัดไป สำหรับจำนวนครั้งของการสุ่มที่แตกต่างกัน			
				5	10	15	30
50	50	0.1	0.1	(2.04,9.56)	(1.37,8.03)	(0.5,8.84)	(1.41,9.13)
100	100	0.2	0.2	(10.80,28.00)	(12.55,28.85)	(11.20,27.46)	(13.14,25.4)
200	200	0.4	0.4	(73.37,95.83)	(69.83,91.97)	(68.51,91.09)	(71.85,86.95)
400	400	0.8	0.8	(319.42,325.38)	(314.35,325.25)	(313.71,326.29)	(313.98,326.02)

จากตารางที่ 4.5 เมื่อพิจารณาช่วงความเชื่อมั่น 95% ของจำนวนของสมาชิกที่ถูกทำเครื่องหมายและสุ่มได้ในการสุ่มตัวอย่างครั้งถัดไป กรณีประชากรขนาด $N = 500$ และขนาดตัวอย่างเหมือนกัน พบว่าถ้าสัดส่วนการสุ่มตัวอย่างเพิ่มขึ้นช่วงของจำนวนสมาชิกที่ถูกทำเครื่องหมายและสุ่มได้ในการสุ่มตัวอย่างครั้งถัดไปจะเข้าใกล้ค่าของจำนวนของสมาชิกที่ทำการสุ่มได้

และถูกทำเครื่องหมายไว้ในการสุ่มตัวอย่างครั้งแรก (M) แต่จะมีค่าใกล้เคียงกันในแต่ละกรณีที่มีสัดส่วนการสุ่มตัวอย่างเท่ากันแต่จำนวนครั้งของการสุ่มตัวอย่างที่แตกต่างกัน

ตารางที่ 4.6 ช่วงความเชื่อมั่น 95% ของจำนวนของสมาชิกที่ถูกทำเครื่องหมายและสุ่มได้ในการสุ่มตัวอย่างครั้งถัดไป กรณีประชากรขนาด $N = 500$ และขนาดตัวอย่างเหมือนกัน

M	n	f1 = M/N	f2 = n/N	ช่วงความเชื่อมั่น 95% ของจำนวนของสมาชิกที่ถูกทำเครื่องหมายและสุ่มได้ในการสุ่มตัวอย่างครั้งถัดไป สำหรับจำนวนครั้งของการสุ่มที่แตกต่างกัน			
				5	10	15	30
50	100	0.1	0.2	(5.31,15.89)	(5.35,13.85)	(5.20,14.26)	(4.55,15.65)
100	50	0.2	0.1	(5.99,12.41)	(4.41,14.39)	(4.08,15.26)	(4.11,13.63)
100	200	0.2	0.4	(26.70,51.70)	(27.99,52.41)	(33.38,45.68)	(32.68,45.86)
200	100	0.4	0.2	(33.39,46.21)	(32.24,49.76)	(27.56,49.78)	(31.79,45.55)

จากตารางที่ 4.6 เมื่อพิจารณาช่วงความเชื่อมั่น 95% ของจำนวนของสมาชิกที่ถูกทำเครื่องหมายและสุ่มได้ในการสุ่มตัวอย่างครั้งถัดไป กรณีประชากรขนาด $N = 500$ และขนาดตัวอย่างแตกต่างกัน พบว่าถ้าสัดส่วนการสุ่มตัวอย่างเพิ่มขึ้นช่วงของจำนวนสมาชิกที่ถูกทำเครื่องหมายและสุ่มได้ในการสุ่มตัวอย่างครั้งถัดไปจะเข้าใกล้ค่าของจำนวนของสมาชิกที่ทำการสุ่มได้และถูกทำเครื่องหมายไว้ในการสุ่มตัวอย่างครั้งแรก (M) แต่จะมีค่าใกล้เคียงกันในแต่ละกรณีที่มีสัดส่วนการสุ่มตัวอย่างเท่ากันแต่จำนวนครั้งของการสุ่มตัวอย่างที่แตกต่างกัน

ตารางที่ 4.7 ช่วงความเชื่อมั่น 95% ของจำนวนของสมาชิกที่ถูกทำเครื่องหมายและสุ่มได้ในการสุ่มตัวอย่างครั้งถัดไป กรณีประชากรขนาด $N = 1,000$ และขนาดตัวอย่างเหมือนกัน

M	n	f1 = M/N	f2 = n/N	ช่วงความเชื่อมั่น 95% ของจำนวนของสมาชิกที่ถูกทำเครื่องหมายและสุ่มได้ในการสุ่มตัวอย่างครั้งถัดไป สำหรับจำนวนครั้งของการสุ่มที่แตกต่างกัน			
				5	10	15	30
100	100	0.1	0.1	(5.28,15.52)	(3.48,16.92)	(4.38,14.02)	(5.58,14.36)
200	200	0.2	0.2	(37.70,49.90)	(28.47,52.73)	(28.35,45.79)	(31.27,50.13)
400	400	0.4	0.4	(147.55,173.65)	(140.79,171.41)	(145.93,167.53)	(146.49,177.31)
800	800	0.8	0.8	(630.65,649.75)	(627.92,653.48)	(627.42,646.32)	(632.39,647.95)

จากตารางที่ 4.7 เมื่อพิจารณาช่วงความเชื่อมั่น 95% ของจำนวนของสมาชิกที่ถูกทำเครื่องหมายและสุ่มได้ในการสุ่มตัวอย่างครั้งถัดไป กรณีประชากรขนาด $N = 1,000$ และขนาด

ตัวอย่างเหมือนกัน พบว่าถ้าสัดส่วนการสุ่มตัวอย่างเพิ่มขึ้นช่วงของจำนวนสมาชิกที่ถูกทำเครื่องหมายและสุ่มได้ในการสุ่มตัวอย่างครั้งถัดไปจะเข้าใกล้ค่าของจำนวนของสมาชิกที่ทำการสุ่มได้ และถูกทำเครื่องหมายไว้ในการสุ่มตัวอย่างครั้งแรก (M) แต่จะมีค่าใกล้เคียงกันในแต่ละกรณีที่มีสัดส่วนการสุ่มตัวอย่างเท่ากันแต่จำนวนครั้งของการสุ่มตัวอย่างที่แตกต่างกัน

ตารางที่ 4.8 ช่วงความเชื่อมั่น 95% ของจำนวนของสมาชิกที่ถูกทำเครื่องหมายและสุ่มได้ในการสุ่มตัวอย่างครั้งถัดไป กรณีประชากรขนาด $N = 1,000$ และขนาดตัวอย่างแตกต่างกัน

M	n	f1 = M/N	f2 = n/N	ช่วงความเชื่อมั่น 95% ของจำนวนของสมาชิกที่ถูกทำเครื่องหมายและสุ่มได้ในการสุ่มตัวอย่างครั้งถัดไป สำหรับจำนวนครั้งของการสุ่มที่แตกต่างกัน			
				5	10	15	30
100	200	0.1	0.2	(12.41,27.17)	(15.99,23.21)	(11.99,27.47)	(11.50,29.1)
200	100	0.2	0.1	(13.81,26.99)	(12.55,29.25)	(14.69,28.65)	(12.48,26.98)
200	400	0.2	0.4	(77.33,91.87)	(68.83,91.17)	(73.05,90.15)	(69.02,88.98)
400	200	0.4	0.2	(69.33,92.27)	(69.69,92.51)	(66.28,92.38)	(68.61,93.65)

จากตารางที่ 4.8 เมื่อพิจารณาช่วงความเชื่อมั่น 95% ของจำนวนของสมาชิกที่ถูกทำเครื่องหมายและสุ่มได้ในการสุ่มตัวอย่างครั้งถัดไป กรณีประชากรขนาด $N = 1,000$ และขนาดตัวอย่างเหมือนกัน พบว่าถ้าสัดส่วนการสุ่มตัวอย่างเพิ่มขึ้นช่วงของจำนวนสมาชิกที่ถูกทำเครื่องหมายและสุ่มได้ในการสุ่มตัวอย่างครั้งถัดไปจะเข้าใกล้ค่าของจำนวนของสมาชิกที่ทำการสุ่มได้ และถูกทำเครื่องหมายไว้ในการสุ่มตัวอย่างครั้งแรก (M) แต่จะมีค่าใกล้เคียงกันในแต่ละกรณีที่มีสัดส่วนการสุ่มตัวอย่างเท่ากันแต่จำนวนครั้งของการสุ่มตัวอย่างที่แตกต่างกัน

บทที่ 5

สรุปผลและข้อเสนอแนะ

5.1 สรุปผลการวิเคราะห์ข้อมูล

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อนำเสนอวิธีการประมาณขนาดของประชากรโดยวิธี Capture – Recapture กรณีที่เป็นการสุ่มตัวอย่างแบบหลายครั้งขึ้นใหม่โดยอาศัยแนวคิดที่พัฒนามาจากวิธีของ Jackson (Modified Jackson Method) และเพื่อเปรียบเทียบวิธีการประมาณดังกล่าวกับวิธีเดิม รวมทั้งเพื่อประมาณค่าช่วงความเชื่อมั่นของจำนวนของสมาชิกที่ถูกทำเครื่องหมายและสุ่มได้ในการสุ่มตัวอย่างครั้งถัดไปที่เหมาะสม โดยในการวิจัยครั้งนี้ได้พัฒนาวิธีการประมาณขนาดของประชากรขึ้นใหม่ทั้งหมด 2 วิธีคือ วิธีที่พัฒนาจาก Jackson โดยใช้ค่ามัธยฐานของเวลา (Modified Jackson using Median Time Method) และวิธีที่พัฒนาจาก Jackson โดยใช้ค่ามัธยฐานของจำนวนของสมาชิกที่ถูกทำเครื่องหมายและสุ่มได้ในการสุ่มตัวอย่างครั้งถัดไป (Modified Jackson using Median m_1 Method) โดยพิจารณาจากค่าเฉลี่ยของเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์เป็นเกณฑ์ในการเปรียบเทียบค่าประมาณจากวิธีการดังกล่าวข้างต้น ทั้ง 3 วิธี ว่าวิธีการใดเหมาะสมที่สุด ซึ่งได้ทำการจำลองค่าประชากร $N = 500$ และ $1,000$ โดยแบ่งออกเป็นกรณีที่มี M และ n มีค่าเท่ากันและมีค่าแตกต่างกัน โดยในแต่ละกรณีทำซ้ำ $10,000$ ครั้ง โดยใช้โปรแกรม Minitab Macros

จากการเปรียบเทียบตัวแบบพยากรณ์ที่ได้จากวิธีการต่าง ๆ ทั้ง 3 วิธี โดยพิจารณาจากค่าเฉลี่ยของเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์แล้ว พบว่าเมื่อพิจารณาค่า MAPE สำหรับจำนวนครั้งของการสุ่มที่แตกต่างกัน ในกรณีประชากรขนาด $N = 500$ และ $1,000$ สำหรับขนาดตัวอย่างเหมือนกันและแตกต่างกัน พบว่าวิธีที่พัฒนาจาก Jackson โดยใช้ค่ามัธยฐานของเวลา จะให้ค่า MAPE ต่ำที่สุด รองลงมาคือวิธีที่พัฒนาจาก Jackson โดยใช้ค่ามัธยฐานของจำนวนของสมาชิกที่ถูกทำเครื่องหมายและสุ่มได้ในการสุ่มตัวอย่างครั้งถัดไป และวิธีที่พัฒนาจาก Jackson ตามลำดับ โดยจะพบว่าถ้าสัดส่วนการสุ่มตัวอย่างและจำนวนครั้งของการสุ่มตัวอย่างเพิ่มขึ้น ค่า MAPE จะมีค่าต่ำลง ยกเว้นวิธีที่พัฒนาจาก Jackson จะพบว่าค่า MAPE สำหรับจำนวนครั้งของการสุ่มที่แตกต่างกันจะมีค่าใกล้เคียงกัน ดังนั้นวิธีการที่เหมาะสมในการประมาณขนาดของประชากรโดยวิธี Capture – Recapture กรณีที่เป็นการสุ่มตัวอย่างแบบหลายครั้งคือวิธีที่พัฒนาจาก Jackson โดยใช้ค่ามัธยฐานของเวลา และเมื่อพิจารณาช่วงความเชื่อมั่น 95% ของจำนวนของสมาชิกที่ถูกทำ

เครื่องหมายและสุ่มได้ในการสุ่มตัวอย่างครั้งถัดไป พบว่าถ้าสัดส่วนการสุ่มตัวอย่างเพิ่มขึ้นช่วงของจำนวนสมาชิกที่ถูกทำเครื่องหมายและสุ่มได้ในการสุ่มตัวอย่างครั้งถัดไปจะเข้าใกล้ค่าของจำนวนของสมาชิกที่ทำการสุ่มได้และถูกทำเครื่องหมายไว้ในในการสุ่มตัวอย่างครั้งแรก (M) แต่จะมีค่าใกล้เคียงกันในแต่ละกรณีที่มีสัดส่วนการสุ่มตัวอย่างเท่ากันแต่จำนวนครั้งของการสุ่มตัวอย่างที่แตกต่างกัน

ดังนั้นจะพบว่า วิธีการที่พัฒนาจาก Jackson ขึ้นใหม่ทั้ง 2 วิธีให้ค่า MAPE ต่ำกว่าวิธีการเดิมในทุกกรณี การประมาณด้วยวิธีการดังกล่าวนี้จึงสามารถนำไปใช้เป็นอีกทางเลือกหนึ่งสำหรับการประมาณขนาดของประชากรในกรณีที่เป็นประชากรแบบเปิด ในกรณีที่เป็นการสุ่มตัวอย่างแบบหลายครั้ง

5.2 ข้อเสนอแนะ

จากการศึกษาวิจัยในครั้งนี้ ผู้วิจัยมีข้อเสนอแนะเพื่อประโยชน์ในการศึกษาวิจัยในอนาคตต่อไปดังนี้

1. ในการศึกษาวิจัยครั้งต่อไป ควรมีการเพิ่มเทคนิคการประมาณขนาดของประชากรอื่นมาประยุกต์ใช้นอกเหนือจากที่กล่าวไว้ในงานวิจัยครั้งนี้ เพื่อจะได้ทำการเปรียบเทียบหาวิธีการประมาณขนาดของประชากรที่เหมาะสมต่อไป

2. ในงานวิจัยนี้ศึกษาเฉพาะกรณีที่เป็นการสุ่มตัวอย่างแบบหลายครั้ง ซึ่งทำเครื่องหมายให้กับสมาชิกแต่ละตัวที่สุ่มได้ในการสุ่มตัวอย่างครั้งแรกเท่านั้น ดังนั้นในการวิจัยครั้งต่อไปควรจะศึกษาในกรณีที่ทำการเครื่องหมายให้กับสมาชิกแต่ละตัวที่สุ่มในการสุ่มตัวอย่างแต่ละครั้งด้วย

บรรณานุกรม

- 1) Aaron D J , Chang Y-F , Markovic N and LaPorte R E (2003), *Estimating the lesbian population: a capture-recapture approach*, Journal of Epidemiology and Community Health 2003; 57:207-209.
- 2) Chapman, D. G. (1951), *Some properties of the hypergeometric distribution with applications to zoological censuses*. University of California Publications Statistics, 1, 131 – 160.
- 3) Fisher N , Turner S W, Pugh R, Taylor C (1994), *Estimated numbers of homeless and homeless mentally ill people in north east Westminster by using capture-recapture analysis*, BMJ 1994; 308:27-30.
- 4) Jackson, C.H.N. (1939), *The Analysis of an Animal Population*, J. Anim. Ecol., 8(2): 238-246.
- 5) Jackson, C.H.N. (1948), *The Analysis of a Tsetse-fly Population III*, Ann. Eugenics, 14(2): 91-108.
- 6) Jolly, G.M. (1965), *Explicit estimates from capture-recapture data with both death and immigration-stochastic model*. Biometrika 52: 225-247.
- 7) Jolly, G.M. (1982), *Mark-recapture models with parameters constant in time*. Biometrics 38-301-321.
- 8) Lincoln, F.C. (1930), *Calculating waterfowl abundance on the basis of banding returns*, U.S. Department of Agricultural Circular, 118, 1-4.
- 9) Mastro, T. D., Kitayaporn, D., Weniger, B.G., Vanichseni, S., Laosunthorn, V., Uneklabh, T., Uneklabh C., Choopanya K., Limpakarnjanarat K. (1994), *Estimating the number of HIV-infected drug users in Bangkok: a capture-recapture method*, American Journal of Public Health, 7, 1094 – 1099.
- 10) Petersen, G. G. J. (1896), *The yearly immigration of young plaice into the limfjord from the german sea*, Report of Danish Biology Statistics, 6, 1 – 48.
- 11) Pezzotti P., Piovesan C., Michieletto F., Zanella F., Rezza G., Gallo G. (2003), *Estimating the cumulative number of human immunodeficiency virus diagnoses by cross-linking from four different sources*, Int J Epidemiol , 32(5) , 778 – 783.
- 12) Sartor F., Walckiers D (2001)., *The prevalence of problematic drug use: Methodological aspects and feasibility in Belgium*, Arch Public Health, 59 , 77-100.

13) Schnabel, Z.E. (1938), *The estimation of the total fish population of a lake*. Amer. Math. Mon. 45 :348-352.

14) Seber, G.A.F. (1982), *The Estimation of animal abundance and related parameters*. Second ed. Macmillan Publishing Co., Inc. New York, NY.

โปรแกรมที่ใช้ในการประมวลผล

โปรแกรมหลัก capture.txt

```
Gmacro
CR_NoBs
Let k1 = 1000
Let k2 = 800
Let k3 = k1-k2
Let k4 = 800
Set c1
k3(0) k2(1)
End
#####Replication R = 10000 times for generate N
Name k90 'n' k91 'M' k92 'I' k93 'R'
Erase c3 - c6
Erase c8 - c22
Do 'I' = 1:10000
Erase c3 - c6
Name c3 'NC' c4 'var_C' c5 'mm' c6 'NSort' c7 't'
Name c8 'NC_Sample' c9 'NS' c10 'beta0' c11 'beta1'
Name k93 'sumbeta0' k94 'sumbeta1' k95 'meanbeta0' k96 'meanbeta1' c12 'mmS'
Name c13 'beta0m' c14 'beta1m' c15 'beta1m1' c16 'beta1m2'
Name c17 'N_Jackson' c18 'N_MedianTime'
Name c19 'N_Median_m' c20 'APE1' c21 'APE2' c22 'APE3'
Noecho
#
Let 'n' = 30
Exec 'recapture.txt' 'n'
Sort 'NC' 'NSort'
#
#Resampling_Regression
#####Simulation with no bootstrapping
##
    Let 'beta1'('I') = [sum('NSort'*'t') - ('n')*mean('t')*mean('NSort')]/[sum('t'*'t') -
('n')*(mean('t')**2)]
    Let 'beta0'('I') = mean('NSort') - 'beta1'('I')*mean('t')
    Let 'N_Jackson'('I') ='beta0'('I')
    Let 'N_MedianTime'('I') ='beta0'('I') + 'beta1'('I')*median('t')
#####Regression using m is independent
    Let 'beta1m1'('I') = [sum('NC'*'mm') - ('n')*mean('mm')*mean('NC')]/
    Let 'beta1m2'('I') = [sum('mm'*'mm') - ('n')*(mean('mm')**2)]
    Let 'beta1m'('I') = ['beta1m1'('I')]/ ['beta1m2'('I')]
    Let 'beta0m'('I') = mean('NC') - 'beta1m'('I')*mean('mm')
    Let 'N_Median_m'('I') ='beta0m'('I') + 'beta1m'('I')*median('mm')
    Let 'APE1'('I') = ABSO(1000 - 'N_Jackson'('I'))*100/1000
    Let 'APE2'('I') = ABSO(1000-'N_MedianTime'('I'))*100/1000
    Let 'APE3'('I') = ABSO(1000 - 'N_Median_m'('I'))*100/1000
Enddo
```

```
Describe 'NC' 'mm' 'beta0' 'beta1' 'beta0m' 'beta1m' 'N_Jackson' 'N_MedianTime'  
Describe 'N_Median_m' 'APE1' 'APE2' 'APE3'  
Endmacro
```

โปรแกรมย่อย recapture.txt

```
Sample k4 c1 c2  
Let k5 = sum(c2)  
Let k6 = round((k2+1)*(k4+1)/(k5+1) - 1)  
Stack k5 'mm' 'mm'  
Stack k6 'NC' 'NC'  
Stack k7 'var_C' 'var_C'
```