ตัวประมาณขนาดประชากรที่ปรับปรุงโดยใช้การถดถอยควอนไทล์ที่ระดับ 0.50

โดย

นางสาวลภัสรดา สิงห์สมบูรณ์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
วิทยาศาสตรมหาบัณฑิต (สถิติประยุกต์)
ภาควิชาคณิตศาสตร์และสถิติ
คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์
ปีการศึกษา 2559

ลิขสิทธิ์ของมหาวิทยาลัยธรรมศาสตร์

Ref. code: 25595809030157UGW

ตัวประมาณขนาดประชากรที่ปรับปรุงโดยใช้
การถดถอยควอนไทล์ที่ระดับ 0.50

โดย

นางสาวลภัสรดา สิงห์สมบูรณ์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร

วิทยาศาสตรมหาบัณฑิต (สถิติประยุกต์)

ภาควิชาคณิตศาสตร์และสถิติ

คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์

ปีการศึกษา 2559

ลิขสิทธิ์ของมหาวิทยาลัยธรรมศาสตร์

Ref. code: 25595809030157UGW

IMPROVED POPULATION SIZE ESTIMATORS USING 0.50 QUANTILE REGRESSION

BY

MISS LAPASRADA SINGHASOMBOON

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF MASTER OF SCIENCE (APPLIED STATISTICS)
DEPARTMENT OF MATHEMATICS AND STATISTICS
FACULTY OF SCIENCE AND TECHNOLOGY
THAMMASAT UNIVERSITY
ACADEMIC YEAR 2016
COPYRIGHT OF THAMMASAT UNIVERSITY

Ref. code: 25595809030157UGW

มหาวิทยาลัยธรรมศาสตร์
คณะวิทยาศาสตร์และเทคโนโลยี

วิทยานิพนธ์

ของ

นางสาวลภัสรดา สิงห์สมบูรณ์

เรื่อง

ตัวประมาณขนาดประชากรที่ปรับปรุงโดยใช้การถดถอยควอนไทล์ที่ระดับ 0.50

ได้รับการตรวจสอบและอนุมัติ ให้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
วิทยาศาสตรมหาบัณฑิต (สถิติประยุกต์)

เมื่อ วันที่ 31 กรกฎาคม พ.ศ. 2560

ประธานกรรมการสอบวิทยานิพนธ์ แสงหล้า ชัยมงคล
(ผู้ช่วยศาสตราจารย์ ดร. แสงหล้า ชัยมงคล)

กรรมการและอาจารย์ที่ปรึกษาวิทยานิพนธ์ ______________________
(ผู้ช่วยศาสตราจารย์ ดร. รัตนา เลิศสุวรรณศรี)

กรรมการสอบวิทยานิพนธ์ ______________________
(อาจารย์ ดร. กฤษณะ ลาน้ำเที่ยง)

กรรมการสอบวิทยานิพนธ์ ______________________
(ผู้ช่วยศาสตราจารย์ ดร. รมิดา ศรีเหรา)

กรรมการสอบวิทยานิพนธ์ ______________________
(อาจารย์ ดร.พัทธ์ชนก ศรีสุรเดชชัย)

คณบดี ______________________
(รองศาสตราจารย์ ปกรณ์ เสริมสุข)

หัวข้อวิทยานิพนธ์	ตัวประมาณขนาดประชากรที่ปรับปรุงโดยใช้การถดถอยควอนไทล์ที่ระดับ 0.50
ชื่อผู้เขียน	นางสาวลภัสรดา สิงห์สมบูรณ์
ชื่อปริญญา	วิทยาศาสตรมหาบัณฑิต (สถิติประยุกต์)
สาขาวิชา/คณะ/มหาวิทยาลัย	สถิติประยุกต์ สาขาวิชาคณิตศาสตร์และสถิติ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์
อาจารย์ที่ปรึกษาวิทยานิพนธ์	ผู้ช่วยศาสตราจารย์ ดร. รัตนา เลิศสุวรรณศรี
ปีการศึกษา	2559

บทคัดย่อ

วิธี Capture-Recapture เป็นวิธีการที่ใช้ในการประมาณขนาดประชากรที่หน่วยตัวอย่างยากต่อการพบเจอ เมื่อไม่นานมานี้ มีผู้วิจัยได้พัฒนาตัวประมาณขนาดประชากร โดยใช้ตัวแบบการถดถอยเชิงเส้นอย่างง่ายแบบถ่วงน้ำหนักบนพื้นฐานของอัตราส่วนของความถี่ที่เกิดขึ้นติดต่อกัน (r_i) ได้แก่ ตัวประมาณขนาดประชากร Generalized Turing ($\hat{N}_{GT}$) และตัวประมาณขนาดประชากร Lanumteang&Böhning ($\hat{N}_{LB}$) งานวิจัยนี้ได้ปรับปรุงตัวประมาณขนาดประชากรขึ้นใหม่ 2 ตัว ($\hat{N}_{NewGT}, \hat{N}_{NewLB}$) ด้วยการใช้ตัวแบบการถดถอยควอนไทล์ที่ระดับ 0.50 บนพื้นฐานของอัตราส่วนของความถี่ที่เกิดขึ้นติดต่อกัน พร้อมทั้งหาความแปรปรวนของตัวประมาณที่ปรับปรุงใหม่ โดยใช้ conditioning technique เพื่อสร้างช่วงเชื่อมั่น 95% ของขนาดประชากร

ผลลัพธ์จากการจำลองสถานการณ์แบบมอนติคาร์โล ได้ว่า ตัวประมาณขนาดประชากร Lanumteang&Böhning ที่ปรับปรุงใหม่ ($\hat{N}_{NewLB}$) เป็นตัวประมาณที่มีประสิทธิภาพที่ให้ค่าความเอนเอียงสัมพัทธ์ต่ำที่สุด และช่วงเชื่อมั่นของขนาดประชากรที่ได้จาก $\hat{N}_{NewLB}$ ให้ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นไม่ต่ำกว่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด ในทุกขนาดประชากรที่ศึกษา นอกจากนี้ผู้วิจัยได้นำตัวประมาณขนาดประชากรที่ได้พัฒนานั้นมาประยุกต์ใช้กับข้อมูลการประมาณจำนวนสปีชีส์ของผีเสื้อมลายู ในเกาะเตียวมัน ประเทศมาเลเซีย โดยตัวประมาณขนาดประชากร $\hat{N}_{NewGT}$ และ $\hat{N}_{NewLB}$ สามารถปรับปรุงประสิทธิภาพของตัวประมาณที่มีอยู่เดิม ($\hat{N}_{GT}$, $\hat{N}_{LB}$) โดยให้ค่าความคลาดเคลื่อนมาตรฐานที่ต่ำกว่า

คำสำคัญ: วิธี Capture-Recapture, การประมาณขนาดประชากร, การแจกแจงทวินามเชิงลบ, การถดถอยควอนไทล์ที่ระดับ 0.50

Ref. code: 25595809030157UGW

Thesis Title	IMPROVED POPULATION SIZE ESTIMATORS USING 0.50 QUANTILE REGRESSION
Author	Miss Lapasrada Singhasomboon
Degree	Master of Science (Applied Statistics)
Major Field/Faculty/University	Mathematics and Statistics Faculty of Science and Technology Thammasat University
Thesis Advisor	Assistant Professor Rattana Lerdsuwansri,Ph.D.
Academic Years	2016

ABSTRACT

Capture-Recapture method is used to estimate the size of an elusive target population. Recently, population size estimators have been introduced based on observed ratios of successive capture frequencies associated with simple weighted regression model (r_i), namely Generalized Turing estimator ($\hat{N}_{GT}$) and Lanumteang&Böhning estimator ($\hat{N}_{LB}$) . In this research, Generalized Turing estimator and Lanumteang&Böhning estimator are improved by using 0.50 quantile regression leading to $\hat{N}_{NewGT}$ and $\hat{N}_{NewLB}$. Additionally, variance of $\hat{N}_{NewGT}$ and $\hat{N}_{NewLB}$ are derived from conditioning technique leading to 95% confident interval for population size *N*.

Results from Monte Carlo simulation showed that $\hat{N}_{NewLB}$ was an efficient estimator giving the lowest relative bias. Confident intervals for population size *N* have satisfying outcomes in which coverage probabilities were fairly closed to the nominal level for all population sizes. Moreover, the improved estimators, $\hat{N}_{NewGT}$ and $\hat{N}_{NewLB}$, were applied to estimate the size of Malayan butterfly from Tioman Island in Malaysia and were compared with the conventional estimators. It was found that the suggested estimators , $\hat{N}_{NewGT}$ and $\hat{N}_{NewLB}$, improved the performance of $\hat{N}_{GT}$ and $\hat{N}_{LB}$ in terms of lower standard error.

keywords: Capture-Recapture, Population size estimation, Negative Binomial, 0.50 Quantile regression

Ref. code: 25595809030157UGW

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงได้ด้วยความกรุณาจากผู้ช่วยศาสตราจารย์ ดร.รัตนา เลิศสุวรรณศรี กรรมการและที่ปรึกษาวิทยานิพนธ์ ที่ได้ให้คำแนะนำและให้คำปรึกษาที่มีประโยชน์ต่องานวิจัย ตลอดจนตรวจทานและแก้ไขข้อบกพร่องต่างๆ จนกระทั่งวิทยานิพนธ์ฉบับนี้เสร็จสมบูรณ์ได้ ผู้วิจัยจึงขอกราบขอบพระคุณอย่างสูงมา ณ ที่นี้ด้วย

ผู้วิจัยกราบขอบพระคุณผู้ช่วยศาสตราจารย์ ดร.แสงหล้า ชัยมงคล ประธานกรรมการสอบวิทยานิพนธ์ที่ได้ให้ข้อเสนอแนะ และข้อคิดเห็นต่างๆ ในการแก้ไขปรับปรุงวิทยานิพนธ์นี้ให้มีความสมบูรณ์มากขึ้น รวมทั้งขอกราบขอบพระคุณผู้ช่วยศาสตราจารย์ ดร.รมิดา ศรีเหรา, อาจารย์ ดร.พัทธ์ชนก ศรีสุรเดชชัย และอาจารย์ ดร.กฤษณะ ลาน้ำเที่ยง คณะกรรมการสอบวิทยานิพนธ์ที่กรุณาเสียสละเวลาช่วยให้ข้อเสนอแนะที่เป็นประโยชน์ต่อการปรับปรุงเนื้อหาวิทยานิพนธ์ฉบับนี้ให้เสร็จสมบูรณ์ยิ่งขึ้น นอกจากนี้ขอขอบพระคุณคณาจารย์ ภาควิชาคณิตศาสตร์และสถิติ คณะวิทยาศาสตร์และเทคโนโลยีที่กรุณามอบวิชาความรู้ให้ และขอบพระคุณเจ้าหน้าที่ประจำโครงการปริญญาโท สาขาสถิติประยุกต์ ที่ช่วยประสานงานและอำนวยความสะดวกให้

สุดท้ายนี้ขอขอบคุณครอบครัวผู้เป็นเสมือนเบื้องหลังแห่งความสำเร็จ ที่คอยสนับสนุนและให้กำลังใจ ตลอดจนให้คำปรึกษาและแรงผลักดันที่สำคัญ ที่ทำให้วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปได้ด้วยดี

นางสาวลภัสรดา สิงห์สมบูรณ์

สารบัญ

	หน้า
บทคัดย่อภาษาไทย	(1)
บทคัดย่อภาษาอังกฤษ	(2)
กิตติกรรมประกาศ	(3)
สารบัญตาราง	(8)
สารบัญภาพ	(11)
บทที่ 1 บทนำ	1
1.1 ความเป็นมาและความสำคัญของปัญหา	1
1.2 วัตถุประสงค์ของการวิจัย	5
1.3 ขอบเขตการศึกษา	5
1.4 เกณฑ์ที่ใช้ในการพิจารณา	6
1.5 ประโยชน์ที่ได้รับจากการศึกษา	9
บทที่ 2 วรรณกรรมและงานวิจัยที่เกี่ยวข้อง	10
2.1 ตัวประมาณขนาดประชากรบนพื้นฐานการแจกแจงทวินามเชิงลบ	10
2.1.1 ตัวประมาณขนาดประชากรของ Generalized Turing	11
2.1.2 ตัวประมาณขนาดประชากรของ Lanumteang&Böhning	14
2.2 คุณสมบัติ Monotonic non-decreasing	15
2.3 Ratio Plot และตัวแบบการถดถอยเชิงเส้น	17
2.4 การถดถอยควอนไทล์	18
2.5 การหาความแปรปรวนของตัวประมาณขนาดประชากร	19

Ref. code: 25595809030157UGW

2.6 วิธีเดลต้า (Delta method) 20
2.7 งานวิจัยที่เกี่ยวข้อง 21

บทที่ 3 วิธีการวิจัย 25

3.1 ตัวประมาณขนาดประชากรของ Generalized Turing
ที่ปรับปรุงโดยใช้การถดถอยควอนไทล์ที่ระดับ 0.50 24
3.2 ตัวประมาณขนาดประชากรของ Lanumteang&Böhning
ที่ปรับปรุงโดยใช้การถดถอยควอนไทล์ที่ระดับ 0.50 25
3.3 การหาความแปรปรวนของตัวประมาณขนาดประชากร 27
3.3.1 การหาความแปรปรวนของ $\hat{N}_{NewGT}$ 27
3.3.2 การหาความแปรปรวนของ $\hat{N}_{NewLB}$ 27
3.3.2.1 หา $Var_n\left\{E(\hat{N}\mid n)\right\}$ 28
3.3.2.2 หา $E_n\left\{Var(\hat{N}\mid n)\right\}$ 28
3.4 การหาช่วงเชื่อมั่นของขนาดประชากร 30
3.4.1 ช่วงเชื่อมั่นของขนาดประชากรโดยตัวประมาณ Generalized Turing 30
3.4.2 ช่วงเชื่อมั่นของขนาดประชากรโดยตัวประมาณ Generalized Turing
ที่ปรับปรุงโดยใช้การถดถอยควอนไทล์ที่ระดับ 0.50 31
3.4.3 ช่วงเชื่อมั่นของขนาดประชากรโดยตัวประมาณ Lanumteang&Böhning 32
3.4.4 ช่วงเชื่อมั่นของขนาดประชากรโดยตัวประมาณ Lanumteang&Böhning
ที่ปรับปรุงโดยใช้การถดถอยควอนไทล์ที่ระดับ 0.50 32
3.5 การดำเนินการวิจัย 33
3.5.1 เปรียบเทียบประสิทธิภาพของตัวประมาณขนาดประชากร
ภายใต้สถานการณ์ต่างๆ 33
3.5.1.1 กำหนดสถานการณ์จำลอง 33
3.5.1.2 ขั้นตอนการดำเนินการวิจัย 34
3.5.2 เปรียบเทียบประสิทธิภาพของช่วงเชื่อมั่นของขนาดประชากร
ภายใต้สถานการณ์ต่างๆ 36

Ref. code: 25595809030157UGW

บทที่ 4 ผลการวิจัยและอภิปรายผล 40

4.1 เปรียบเทียบประสิทธิภาพของตัวประมาณขนาดประชากร 41
4.1.1 การจำลองภายใต้ตัวแบบการแจกแจงปัวซง 41
4.1.2 การจำลองภายใต้ตัวแบบการแจกแจงปัวซงแบบผสมที่มีพารามิเตอร์ λ_1, λ_2 44
4.1.3 การจำลองภายใต้ตัวแบบการแจกแจงทวินามเชิงลบ 45
4.2 เปรียบเทียบประสิทธิภาพของช่วงเชื่อมั่นของขนาดประชากร 57
4.2.1 การจำลองภายใต้ตัวแบบการแจกแจงปัวซง 57
4.2.2 การจำลองภายใต้ตัวแบบการแจกแจงปัวซงแบบผสมที่มีพารามิเตอร์ λ_1, λ_2 59
4.2.3 การจำลองภายใต้ตัวแบบการแจกแจงทวินามเชิงลบ 61
4.3 การประยุกต์ใช้กับข้อมูลจริง 72

บทที่ 5 สรุปผลการวิจัยและข้อเสนอแนะ 75

5.1 ผลการศึกษาจากค่าความเอนเอียงสัมพัทธ์ $(Rbias)$ 76
5.2 ผลการศึกษาจากค่าความแปรปรวนสัมพัทธ์ $(RVar)$ 78
5.3 ผลการศึกษาจากค่าความคลาดเคลื่อนกำลังสองสัมพัทธ์ $(RMSE)$ 79
5.4 ผลการศึกษาจากค่าประมาณสัมประสิทธิ์ความเชื่อมั่นของขนาดประชากร $(1-\hat{\alpha})$ 81
5.5 ผลการศึกษาจากค่าประมาณความกว้างเฉลี่ยของช่วงเชื่อมั่น (Average Length) 83
5.6 ผลการศึกษาเปรียบเทียบประสิทธิภาพของตัวประมาณขนาดประชากร 85
5.7 ผลการศึกษาเปรียบเทียบประสิทธิภาพของช่วงเชื่อมั่นของขนาดประชากร 86
5.8 ข้อเสนอแนะ 87
5.8.1 ข้อเสนอแนะเกี่ยวกับการวิจัย 87
5.8.2 ข้อเสนอแนะเกี่ยวกับการนำไปประยุกต์ใช้ 87

รายการอ้างอิง 88

ภาคผนวก

ภาคผนวก ก โปรแกรมที่ใช้ในการศึกษา 91

Ref. code: 25595809030157UGW

ประวัติผู้เขียน 122

Ref. code: 25595809030157UGW

สารบัญตาราง

ตารางที่ หน้า

1.1 แสดงจำนวนครั้งและจำนวนหน่วยตัวอย่างที่ถูกพบ i ครั้ง 1

4.1 แสดงค่าประมาณขนาดประชากร ค่าความเอนเอียงสัมพัทธ์ ค่าความแปรปรวนสัมพัทธ์และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ของตัวประมาณทั้ง 4 ตัว ภายใต้การแจกแจงปัวซง $p_i \sim Poi(\lambda)$ เมื่อกำหนดให้ $\lambda = 3$ และขนาดประชากร $N = 500, 1{,}000, 5{,}000, 10{,}000, 50{,}000$ และ $100{,}000$ 48

4.2 แสดงค่าประมาณขนาดประชากร ค่าความเอนเอียงสัมพัทธ์ ค่าความแปรปรวนสัมพัทธ์และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ของตัวประมาณทั้ง 4 ตัว ภายใต้การแจกแจงปัวซง $p_i \sim Poi(\lambda)$ เมื่อกำหนดให้ $\lambda = 5$ และขนาดประชากร $N = 500, 1{,}000, 5{,}000, 10{,}000, 50{,}000$ และ $100{,}000$ 49

4.3 แสดงค่าประมาณขนาดประชากร ค่าความเอนเอียงสัมพัทธ์ ค่าความแปรปรวนสัมพัทธ์และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ของตัวประมาณทั้ง 4 ตัว ภายใต้การแจกแจงปัวซง $p_i \sim Poi(\lambda)$ เมื่อกำหนดให้ $\lambda = 7$ และขนาดประชากร $N = 500, 1{,}000, 5{,}000, 10{,}000, 50{,}000$ และ $100{,}000$ 50

4.4 แสดงค่าประมาณขนาดประชากร ค่าความเอนเอียงสัมพัทธ์ ค่าความแปรปรวนสัมพัทธ์และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ของตัวประมาณทั้ง 4 ตัว ภายใต้การแจกแจงปัวซงแบบผสมที่มีพารามิเตอร์ λ_1, λ_2 หรือ $p_i \sim 0.5Poi(\lambda_1) + 0.5Poi(\lambda_2)$ โดยที่ $(\lambda_1, \lambda_2) = (3,5)$ และขนาดประชากร $N = 500, 1{,}000, 5{,}000, 10{,}000, 50{,}000, 100{,}000$ 51

4.5 แสดงค่าประมาณขนาดประชากร ค่าความเอนเอียงสัมพัทธ์ ค่าความแปรปรวนสัมพัทธ์และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ของตัวประมาณทั้ง 4 ตัว ภายใต้การแจกแจงปัวซงแบบผสมที่มีพารามิเตอร์ λ_1, λ_2 หรือ $p_i \sim 0.5Poi(\lambda_1) + 0.5Poi(\lambda_2)$ โดยที่ $(\lambda_1, \lambda_2) = (7,5)$ และขนาดประชากร $N = 500, 1{,}000, 5{,}000, 10{,}000, 50{,}000, 100{,}000$ 52

4.6 แสดงค่าประมาณขนาดประชากร ค่าความเอนเอียงสัมพัทธ์ ค่าความแปรปรวนสัมพัทธ์และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ของตัวประมาณทั้ง 4 ตัว ภายใต้การแจกแจงทวินามเชิงลบ ที่มีพารามิเตอร์ π, κ, $p_i \sim NB(\kappa, \pi)$ โดยที่ $\kappa = 4, \pi = 0.5$ และขนาดประชากร $N = 500, 1{,}000, 5{,}000, 10{,}000, 50{,}000, 100{,}000$ 53

Ref. code: 25595809030157UGW

4.7 แสดงค่าประมาณขนาดประชากร ค่าความเอนเอียงสัมพัทธ์ ค่าความแปรปรวนสัมพัทธ์และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ของตัวประมาณทั้ง 4 ตัว ภายใต้การแจกแจงทวินามเชิงลบ ที่มีพารามิเตอร์ $\pi,\kappa,\ \ p_i \sim NB(\kappa,\pi)$ โดยที่ $\kappa=4,\pi=0.8$ และขนาดประชากร $N=500, 1{,}000, 5{,}000, 10{,}000, 50{,}000, 100{,}000$ 54

4.8 แสดงค่าประมาณขนาดประชากร ค่าความเอนเอียงสัมพัทธ์ ค่าความแปรปรวนสัมพัทธ์และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ของตัวประมาณทั้ง 4 ตัว ภายใต้การแจกแจงทวินามเชิงลบ ที่มีพารามิเตอร์ $\pi,\kappa,\ \ p_i \sim NB(\kappa,\pi)$ โดยที่ $\kappa=9,\pi=0.5$ และขนาดประชากร $N=500, 1{,}000, 5{,}000, 10{,}000, 50{,}000, 100{,}000$ 55

4.9 แสดงค่าประมาณขนาดประชากร ค่าความเอนเอียงสัมพัทธ์ ค่าความแปรปรวนสัมพัทธ์และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ของตัวประมาณทั้ง 4 ตัว ภายใต้การแจกแจงทวินามเชิงลบ ที่มีพารามิเตอร์ $\pi,\kappa,\ \ p_i \sim NB(\kappa,\pi)$ โดยที่ $\kappa=9,\pi=0.8$ และขนาดประชากร $N=500, 1{,}000, 5{,}000, 10{,}000, 50{,}000, 100{,}000$ 56

4.10 แสดงค่าประมาณสัมประสิทธิ์ความเชื่อมั่นและค่าประมาณความกว้างเฉลี่ยของช่วงเชื่อมั่นของตัวประมาณทั้ง 4 ภายใต้การแจกแจงปัวซง $p_i \sim Poi(\lambda)$ เมื่อกำหนดให้ $\lambda=3$ และขนาดประชากร $N=500, 1{,}000, 5{,}000, 10{,}000, 50{,}000, 100{,}000$ 63

4.11 แสดงค่าประมาณสัมประสิทธิ์ความเชื่อมั่นและค่าประมาณความกว้างเฉลี่ยของช่วงเชื่อมั่นของตัวประมาณทั้ง 4 ภายใต้การแจกแจงปัวซง $p_i \sim Poi(\lambda)$ เมื่อกำหนดให้ $\lambda=5$ และขนาดประชากร $N=500, 1{,}000, 5{,}000, 10{,}000, 50{,}000, 100{,}000$ 64

4.12 แสดงค่าประมาณสัมประสิทธิ์ความเชื่อมั่นและค่าประมาณความกว้างเฉลี่ยของช่วงเชื่อมั่นของตัวประมาณทั้ง 4 ภายใต้การแจกแจงปัวซง $p_i \sim Poi(\lambda)$ เมื่อกำหนดให้ $\lambda=7$ และขนาดประชากร $N=500, 1{,}000, 5{,}000, 10{,}000, 50{,}000, 100{,}000$ 65

4.13 แสดงค่าประมาณสัมประสิทธิ์ความเชื่อมั่นและค่าประมาณความกว้างเฉลี่ยของช่วงเชื่อมั่นของตัวประมาณทั้ง 4 ภายใต้การแจกแจงปัวซงแบบผสมที่มีพารามิเตอร์ λ_1,λ_2 หรือ $p_i \sim 0.5Poi(\lambda_1)+0.5Poi(\lambda_2)$ โดยที่ $(\lambda_1,\lambda_2)=(3,5)$ และ $N=500, 1{,}000, 5{,}000, 10{,}000$ $50{,}000, 100{,}000$ 66

4.14 แสดงค่าประมาณสัมประสิทธิ์ความเชื่อมั่นและค่าประมาณความกว้างเฉลี่ยของช่วงเชื่อมั่นของตัวประมาณทั้ง 4 ภายใต้การแจกแจงปัวซงแบบผสมที่มีพารามิเตอร์ λ_1,λ_2 หรือ $p_i \sim 0.5Poi(\lambda_1)+0.5Poi(\lambda_2)$ โดยที่ $(\lambda_1,\lambda_2)=(7,5)$ และ $N=500, 1{,}000, 5{,}000, 10{,}000$ $50{,}000, 100{,}000$ 67

Ref. code: 25595809030157UGW

4.15 แสดงค่าประมาณสัมประสิทธิ์ความเชื่อมั่นและค่าประมาณความกว้างเฉลี่ยของช่วงเชื่อมั่นของตัวประมาณทั้ง 4 ภายใต้การแจกแจงทวินามเชิงลบ ที่มีพารามิเตอร์ $\pi,\kappa,\ p_i \sim NB(\kappa,\pi)$ โดยที่ $\kappa=4,\pi=0.5$ และขนาดประชากร $N=500, 1{,}000, 5{,}000, 10{,}000, 50{,}000, 100{,}000$ 68

4.16 แสดงค่าประมาณสัมประสิทธิ์ความเชื่อมั่นและค่าประมาณความกว้างเฉลี่ยของช่วงเชื่อมั่นของตัวประมาณทั้ง 4 ภายใต้การแจกแจงทวินามเชิงลบ ที่มีพารามิเตอร์ $\pi,\kappa,\ p_i \sim NB(\kappa,\pi)$ โดยที่ $\kappa=4,\pi=0.8$ และขนาดประชากร $N=500, 1{,}000, 5{,}000, 10{,}000, 50{,}000, 100{,}000$ 69

4.17 แสดงค่าประมาณสัมประสิทธิ์ความเชื่อมั่นและค่าประมาณความกว้างเฉลี่ยของช่วงเชื่อมั่นของตัวประมาณทั้ง 4 ภายใต้การแจกแจงทวินามเชิงลบ ที่มีพารามิเตอร์ $\pi,\kappa,\ p_i \sim NB(\kappa,\pi)$ โดยที่ $\kappa=9,\pi=0.5$ และขนาดประชากร $N=500, 1{,}000, 5{,}000, 10{,}000, 50{,}000, 100{,}000$ 70

4.18 แสดงค่าประมาณสัมประสิทธิ์ความเชื่อมั่นและค่าประมาณความกว้างเฉลี่ยของช่วงเชื่อมั่นของตัวประมาณทั้ง 4 ภายใต้การแจกแจงทวินามเชิงลบ ที่มีพารามิเตอร์ $\pi,\kappa,\ p_i \sim NB(\kappa,\pi)$ โดยที่ $\kappa=9,\pi=0.8$ และขนาดประชากร $N=500, 1{,}000, 5{,}000, 10{,}000, 50{,}000, 100{,}000$ 71

4.19 จำนวนครั้งและจำนวนความถี่ที่พบจำนวนสปีชีส์ของผีเสื้อมลายู ในเกาะเตียวมัน ประเทศมาเลเซีย ซึ่งจำนวนของจำนวนสปีชีส์ของผีเสื้อมลายู รวมทั้งสิ้น 620 สปีชีส์ 72

4.20 แสดงค่าประมาณขนาดประชากร, ค่าคลาดเคลื่อนมาตรฐานของตัวประมาณขนาดประชากร และช่วงเชื่อมั่น 95% ของขนาดประชากรสปีชีส์ของผีเสื้อมลายู 74

5.1 แสดงผลสรุปตัวประมาณขนาดประชากรที่มีประสิทธิภาพสูงสุดเมื่อพิจารณาค่าความเอนเอียงสัมพัทธ์ ($Rbias$) 76

5.2 แสดงผลสรุปตัวประมาณขนาดประชากรที่มีประสิทธิภาพสูงสุดเมื่อพิจารณาค่าความแปรปรวนสัมพัทธ์ ($RVar$) 78

5.3 แสดงผลสรุปตัวประมาณขนาดประชากรที่มีประสิทธิภาพสูงสุด เมื่อพิจารณาค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ ($RMSE$) 79

5.4 แสดงผลสรุปช่วงเชื่อมั่นของตัวประมาณขนาดประชากรที่มีประสิทธิสูงสุดเมื่อพิจารณาค่าประมาณสัมประสิทธิ์ความเชื่อมั่นของขนาดประชากร ($1-\hat{\alpha}$) 81

5.5 แสดงผลสรุปช่วงเชื่อมั่นของตัวประมาณขนาดประชากรที่มีประสิทธิสูงสุดเมื่อพิจารณาค่าประมาณความกว้างเฉลี่ยของช่วงเชื่อมั่น (Average Length) 83

Ref. code: 25595809030157UGW

สารบัญภาพ

ภาพที่ หน้า

1.1 แผนภาพการกระจายของตัวแปร x กับ y โดยเปรียบเทียบระหว่างการถดถอยเข้าสู่ค่าเฉลี่ยกับการถดถอยเข้าสู่ค่ามัธยฐาน 4

3.1 เปรียบเทียบประสิทธิภาพของตัวประมาณขนาดประชากร 38

3.2 เปรียบเทียบประสิทธิภาพช่วงเชื่อมั่นของขนาดประชากร 39

4.1 (ก) กราฟแสดงความสัมพันธ์ระหว่าง r_x กับ x ของการถดถอยเชิงเส้นแบบถ่วงน้ำหนักและการถดถอยควอนไทล์ที่ $\tau = 0.50$ 73

4.1 (ข) กราฟแสดงความสัมพันธ์ระหว่าง $\log(r_x)$ กับ x ของการถดถอยเชิงเส้นแบบถ่วงน้ำหนักและการถดถอยควอนไทล์ที่ $\tau = 0.50$ 73

Ref. code: 25595809030157UGW

บทที่ 1
บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

วิธี Capture-Recapture มีวัตถุประสงค์เพื่อใช้ในการประมาณขนาดของประชากรที่หน่วยตัวอย่างมีลักษณะยากต่อการพบเจอโดยทั่วไป ซึ่งเครื่องมือที่ใช้ในการเก็บรวบรวมข้อมูล เช่น การวางกับดัก การจดบันทึกและการลงทะเบียน เป็นต้น เดิมวิธี Capture-Recapture เป็นวิธีการที่ใช้ในการประมาณขนาดของประชากรสัตว์ป่าในพื้นที่กว้าง โดยที่ประชากรสัตว์มีการอพยพย้ายถิ่นฐานอยู่ตลอดเวลา ทำให้ยากต่อการเก็บรวบรวมข้อมูลจากทุกหน่วยของประชากร ตัวอย่างเช่น ต้องการประมาณขนาดประชากรหมีกรีซลีย์เพศเมียในอุทยานแห่งชาติเยลโลว์สโตน (Chao & Huggins, 2005) ซึ่งโดยปกติแล้วไม่สามารถที่จะเก็บรวบรวมข้อมูลจากทุกหน่วยของประชากรได้ จึงต้องอาศัยวิธี Capture-Recapture นี้ช่วยในการประมาณขนาดของประชากร ต่อมาวิธี Capture-Recapture ได้ถูกนำมาประยุกต์ใช้ในการประมาณขนาดประชากรของมนุษย์ และใช้อย่างแพร่หลายในงานด้านต่างๆ เช่น ด้านสังคมศาสตร์ ด้านสัตวแพทยศาสตร์ ด้านสาธารณสุขศาสตร์ ด้านระบาดวิทยาและด้านอื่นๆ ตัวอย่างเช่น การประมาณจำนวนผู้ป่วยโรคมาลาเรียในประเทศเนเธอร์แลนด์ (Lanumteang,2010) การประมาณจำนวนแรงงานบังคับทั่วโลกขององค์การแรงงานระหว่างประเทศหรือไอโอแอล (Bales,2012) และการประมาณจำนวนผู้ไร้ที่อยู่อาศัยในเมืองอูเทรคต์ ประเทศเนเธอร์แลนด์ (van der Heijden, Cruyff and Böhning,2014a) เป็นต้น

ข้อมูลที่ได้จากการเก็บรวบรวมด้วยวิธี Capture-Recapture นี้มีลักษณะเป็นข้อมูลแจงนับ (Count data) แสดงรายละเอียดได้ดังตารางที่ 1.1

ตารางที่ 1.1 แสดงจำนวนครั้งและจำนวนหน่วยตัวอย่างที่ถูกพบ i ครั้ง

จำนวนครั้งที่หน่วยตัวอย่างถูกพบ (i)	0	1	2	3	...	m
จำนวนหน่วยตัวอย่างที่ถูกพบ (f_i)	-	f_1	f_2	f_3	...	f_m

โดยกำหนดให้ N คือ ขนาดประชากร

f_i คือ จำนวนหน่วยตัวอย่างที่ถูกพบ i ครั้ง ; $i=0,1,2,3,...,m$

เมื่อ m คือ จำนวนครั้งสูงสุดที่พบ

f_0 คือ จำนวนหน่วยตัวอย่างที่ไม่ถูกพบ

Ref. code: 25595809030157UGW

n คือ จำนวนหน่วยตัวอย่างที่ถูกพบหรือ $n = \sum_{i=1}^{m} f_i$

เห็นได้ว่าขนาดประชากร N หาได้จากผลรวมของจำนวนหน่วยตัวอย่างที่ถูกพบ (n) กับจำนวนหน่วยตัวอย่างไม่ถูกพบ (f_0) หรือเขียนได้ว่า

$$N = n + f_0$$

เนื่องจากหน่วยตัวอย่างที่ไม่ถูกพบนั้นไม่เคยปรากฏ เราจึงไม่ทราบค่า f_0 และจำเป็นต้องประมาณเพื่อให้ได้ค่าประมาณของขนาดประชากร เมื่อพิจารณาตัวแปรสุ่ม X ซึ่งแทน จำนวนครั้งที่เกิดเหตุการณ์ในช่วงเวลาที่ศึกษาของวิธี Capture-Recapture และกำหนดให้ f_x แทนจำนวนหน่วยตัวอย่างที่ถูกพบ x ครั้ง โดยที่ $x = 0,1,2,3,...,m$ จากค่าประมาณขนาดประชากรคือ $\hat{N} = n + \hat{f}_0$ จะอาศัยความน่าจะเป็นมาเกี่ยวข้อง โดยพิจารณาจากลักษณะข้อมูลที่มีลักษณะการแจงนับ ซึ่งการแจกแจงที่เป็นที่นิยมได้แก่ การแจกแจงปัวซง (Poisson distribution) ที่มีพารามิเตอร์ λ และกำหนดให้ p_x คือความน่าจะเป็นที่หน่วยตัวอย่างจะถูกพบ x ครั้ง ซึ่งจากความสัมพันธ์ $p_x = \frac{f_x}{N}$ ดังนั้นหา f_x ได้จาก $f_x = Np_x$ ดังนั้น สามารถเขียนสมการการหาค่าประมาณขนาดประชากรได้อีกรูปแบบหนึ่ง คือ

$$\hat{N} = \frac{n}{1 - \hat{p}_0}$$

สำหรับตัวแปรสุ่มที่มีการแจกแจงปัวซง (Poisson distribution) หากพิจารณาค่าพารามิเตอร์ λ เมื่อพารามิเตอร์ λ มีค่าคงที่หรือความน่าจะเป็นที่หน่วยตัวอย่างแต่ละหน่วยจะถูกพบ x ครั้ง มีค่าเท่ากัน เราเรียกกรณีนี้ว่า Homogeneous Poisson Model ซึ่งมีฟังก์ชันความน่าจะเป็นของตัวแปรสุ่มคือ $P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!} \quad ;\lambda > 0$ โดยที่ตัวประมาณขนาดของประชากรที่พัฒนาภายใต้ Homogeneous Poisson Model เช่น ตัวประมาณขนาดประชากรด้วยวิธี Maximum Likelihood และตัวประมาณขนาดประชากร Turing (Good, 1953) เป็นต้น และเมื่อพารามิเตอร์ λ มีค่าไม่คงที่หรือความน่าจะเป็นที่หน่วยตัวอย่างแต่ละหน่วยถูกพบ x ครั้ง มีค่าไม่เท่ากัน ซึ่งขึ้นอยู่กับการแจกแจงของพารามิเตอร์ λ เราเรียกกรณีนี้ว่า Heterogeneous Poisson Model ฟังก์ชันความหนาแน่นความน่าจะเป็นของตัวแปรสุ่ม คือ $P(X = x) = \int_0^\infty \frac{e^{-\lambda}\lambda^x}{x!} g(\lambda)d\lambda$ โดยที่ $g(\lambda)$ คือฟังก์ชันหนาแน่นความน่าจะเป็น (Probability density function) ของตัวแปรสุ่ม Λ ซึ่งตัวประมาณขนาดของประชากรที่พัฒนาภายใต้ Heterogeneous Poisson Model เช่น ตัวประมาณขนาดประชากรของ Chao (Chao,1987), ตัวประมาณขนาดประชากรของ Lanumteang&Böhning (Lanumteang&Böhning,2011) และตัวประมาณขนาดประชากร Generalized Turing

(Böhning et al.,2013) เป็นต้น

สำหรับ Heterogeneous Poisson Model กรณีที่ $g(\lambda)$ เป็นฟังก์ชันความหนาแน่นความน่าจะเป็นของการแจกแจงแกมมา (Gamma distribution) ที่มี Shape parameter κ และ Scale parameter $\theta = \frac{1-\pi}{\pi}$ แล้ว จะได้

$$P(X = x) = \int_0^{\infty} \frac{e^{-\lambda}\lambda^x}{x!} g(\lambda) d\lambda$$

$$= \frac{\Gamma(\kappa + x)}{\Gamma(x+1)\Gamma(\kappa)} \pi^{\kappa}(1-\pi)^x \quad , x = 0,1,2,...$$

ซึ่งจากฟังก์ชันความน่าจะเป็นข้างต้นคือ การแจกแจงทวินามเชิงลบ (Negative Binomial distribution) ที่มีพารามิเตอร์ κ และ π โดยตัวประมาณขนาดประชากรที่พัฒนาภายใต้การแจกแจงทวินามเชิงลบ เช่น ตัวประมาณขนาดประชากร Generalized Turing (Böhning et al.,2013) และตัวประมาณขนาดประชากรของ Lanumteang&Böhning (Lanumteang&Böhning,2011) เป็นต้น

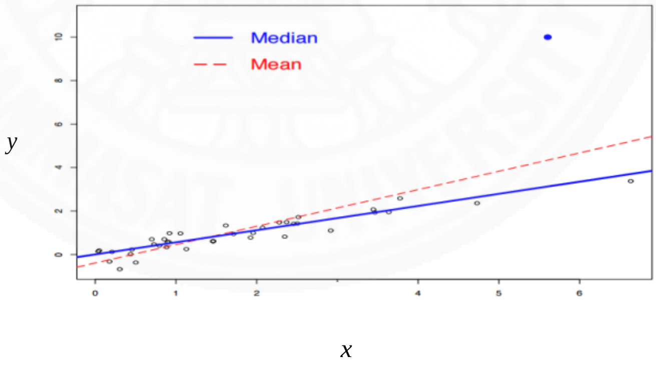
ตัวประมาณขนาดประชากรที่พัฒนาภายใต้ Heterogeneous Poisson Model เริ่มต้นมาจากในปี ค.ศ. 1987 Chao ได้พัฒนาตัวประมาณขนาดประชากรภายใต้ Heterogeneous Poisson Model ขึ้น ที่เรียกว่าตัวประมาณขนาดประชากรของ Chao โดยมีฟังก์ชันความน่าจะเป็นของตัวแปรสุ่มคือ $P(X = x) = \int_0^{\infty} \frac{e^{-\lambda}\lambda^x}{x!} g(\lambda) d\lambda$ ซึ่งหลักการของ Chao ใช้อสมการของ Cauchy-Schwartz ในการจัดรูป ซึ่งพบว่ามีคุณสมบัติ monotonic non-decreasing กล่าวคือ $\frac{p_1}{p_0} < \frac{2p_2}{p_1} < \frac{3p_3}{p_2} < ...$ หรืออยู่ในรูป $\frac{(x+1)p_{x+1}}{p_x}$ และเมื่อกำหนดให้ $r_x = \frac{(x+1)p_{x+1}}{p_x}$ แสดงให้เห็นว่า r_x จะมีคุณสมบัติ monotonic non-decreasing

ต่อมาในปี ค.ศ. 2013 Böhning และคณะ ได้นำเสนอเครื่องมือที่เรียกว่า Ratio Plot เพื่อใช้ตรวจสอบข้อมูลแจงนับว่าเป็น Homogeneous Poisson Model หรือ Heterogeneous Poisson Model โดยการพล็อตกราฟความสัมพันธ์ระหว่าง $\hat{r}_x = \frac{(x+1)f_{x+1}}{f_x}$ กับ x ในกรณีที่ข้อมูลแจงนับมีลักษณะ Homogeneous Poisson Model กราฟ Ratio Plot จะแสดงเส้นตรงที่มีความชันเป็นศูนย์ และกรณีที่ข้อมูลแจงนับมีลักษณะ Heterogeneous Poisson Model กราฟ Ratio Plot จะแสดงคุณสมบัติ monotonic non-decreasing หรือแสดงเส้นตรงที่มีความชันเป็นบวก นอกจากนี้ Böhning และคณะ ยังได้นำเสนอกราฟ Ratio Plot ภายใต้การแจกแจงทวินามเชิงลบที่มีพารามิเตอร์

Ref. code: 25595809030157UGW

κ และ π จาก $p_x = \frac{\Gamma(\kappa + x)}{\Gamma(x+1)\Gamma(\kappa)} \pi^{\kappa}(1-\pi)^x$ และจาก $r_x = \frac{(x+1)p_{x+1}}{p_x}$ จัดรูปจะได้ $r_x = (1-\pi)\kappa + (1-\pi)x$ ซึ่งเขียนอยู่ในรูปของสมการเส้นตรงแสดงความสัมพันธ์ระหว่าง r_x กับ x ซึ่งกราฟแสดงความชันเป็นบวก นั่นคือลักษณะ Heterogeneous Poisson Model

ตัวประมาณขนาดประชากร Generalized Turing $(\hat{N}_{GT})$ และตัวประมาณขนาดประชากรของ Lanumteang&Böhning $(\hat{N}_{LB})$ อาศัย Ratio Plot และการถดถอยเชิงเส้นอย่างง่ายแบบถ่วงน้ำหนัก (Weighted Least Square Method) ของ $\hat{r}_x$ กับ x เป็นเครื่องมือเพื่อประมาณค่าพารามิเตอร์ของตัวแบบความน่าจะเป็น เพื่อที่จะได้ $\hat{f}_0$ และ $\hat{N}$ ตามลำดับ อย่างไรก็ตาม พบว่าตัวประมาณขนาดประชากร Generalized Turing มีประสิทธิภาพต่ำในกรณีที่ขนาดประชากรมีขนาดเล็ก นอกจากนี้ตัวประมาณขนาดประชากร Lanumteang&Böhning มีค่าความแปรปรวนสูงกว่าตัวประมาณอื่นๆ ซึ่งจากการทบทวนวรรณกรรมของการถดถอยเชิงเส้นอย่างง่ายแบบถ่วงน้ำหนักซึ่งเป็นการประมาณค่าพารามิเตอร์โดยถดถอยเข้าสู่ค่าเฉลี่ย และการถดถอย ควอนไทล์ใช้ในการประมาณค่าพารามิเตอร์ในตัวแบบโดยขึ้นอยู่กับการกำหนดระดับของ τ โดยที่ τ คือ ระดับควอนไทล์ที่มีค่าอยู่ระหว่าง 0 ถึง 1 , $\tau \in (0,1)$ งานวิจัยนี้ เรากำหนดให้ $\tau = 0.50$ นั่นคือ เป็นการประมาณค่าพารามิเตอร์โดยเป็นการถดถอยเข้าสู่ค่ามัธยฐาน ซึ่งค่ามัธยฐานนั้นมีความแกร่งเมื่อข้อมูลมีค่าผิดปกติ (Outlier) ดังจะเห็นได้จากภาพที่ 1.1 (Koenker and Bassett,1978)



ภาพที่ 1.1 แผนภาพการกระจายของตัวแปร x กับ y โดยเปรียบเทียบระหว่างการถดถอยเข้าสู่ค่าเฉลี่ยกับการถดถอยเข้าสู่ค่ามัธยฐาน

จากภาพที่ 1.1 แสดงแผนภาพการกระจายของตัวแปร x กับ y โดยเปรียบเทียบระหว่างการถดถอยเข้าสู่ค่าเฉลี่ยกับการถดถอยเข้าสู่ค่ามัธยฐาน จะเห็นได้ว่า ค่าสังเกตแต่ละค่ามีค่าใกล้เคียง

กับเส้นถดถอยของการถดถอยเข้าสู่ค่ามัยธฐานมากกว่าเส้นถดถอยของการถดถอยเข้าสู่ค่าเฉลี่ย ด้วยเหตุนี้เราจึงต้องการปรับปรุงตัวประมาณขนาดประชากร Generalized Turing และตัวประมาณขนาดประชากร Lanumteang&Böhning โดยใช้การถดถอยควอนไทล์ที่ระดับ 0.50

นอกจากพัฒนาตัวประมาณขนาดประชากรขึ้นใหม่เพื่อปรับปรุงประสิทธิภาพของ $\hat{N}_{GT}$ และ $\hat{N}_{LB}$ ซึ่งเป็นตัวประมาณเดิมที่มีอยู่แล้ว เรายังสนใจการหาความแปรปรวนของตัวประมาณขนาดประชากรที่พัฒนาขึ้นใหม่ เพื่อใช้ประโยชน์ในการสร้างช่วงเชื่อมั่นของขนาดประชากร

1.2 วัตถุประสงค์ของการวิจัย

1.2.1 เพื่อปรับปรุงตัวประมาณขนาดประชากร Generalized Turing และตัวประมาณขนาดประชากร Lanumteang&Böhning โดยใช้การถดถอยควอนไทล์ที่ระดับ 0.50

1.2.2 เพื่อหาค่าประมาณความแปรปรวนของตัวประมาณขนาดประชากรที่ปรับปรุงขึ้นใหม่ที่ได้จากข้อ (1.2.1)

1.2.3 เพื่อศึกษาประสิทธิภาพของตัวประมาณขนาดประชากร $\hat{N}_{GT}$, $\hat{N}_{NewGT}$, $\hat{N}_{LB}$ และ $\hat{N}_{NewLB}$ ภายใต้สถานการณ์ต่างๆ

1.2.4 เพื่อศึกษาประสิทธิภาพของช่วงเชื่อมั่นของขนาดประชากรที่สร้างจาก $\hat{N}_{GT}$, $\hat{N}_{NewGT}$, $\hat{N}_{LB}$ และ $\hat{N}_{NewLB}$ ภายใต้สถานการณ์ต่างๆ

1.3 ขอบเขตการศึกษา

ในงานวิจัยนี้เป็นการศึกษาโดยการจำลองข้อมูลและประมวลผลโดยใช้โปรแกรม R Studio มีขอบเขตการศึกษาดังนี้

1.3.1 กำหนดขนาดประชากร (N) ที่ศึกษาดังนี้ กรณีประชากรขนาดเล็ก $N = 500,\ 1{,}000$, ประชากรขนาดกลาง $N = 5{,}000,\ 10{,}000$ และประชากรขนาดใหญ่ $N = 50{,}000,\ 100{,}000$

1.3.2 กำหนดการแจกแจงที่ศึกษา ได้แก่

1.3.2.1 การแจกแจงปัวซง (Poisson distribution) ที่มีพารามิเตอร์ λ โดยกำหนดให้ $\lambda \in \{3,5,7\}$

1.3.2.2 การแจกแจงปัวซงแบบผสมที่มีพารามิเตอร์ λ_1, λ_2 โดยกำหนดให้พารามิเตอร์ $(\lambda_1, \lambda_2) = (3,5)$ และ $(\lambda_1, \lambda_2) = (7,5)$

Ref. code: 25595809030157UGW

1.3.2.3 การแจกแจงทวินามเชิงลบ (Negative Binomial distribution) ที่มีพารามิเตอร์ π,κ โดยกำหนดให้ $\pi \in \{0.50, 0.80\}$ และ $\kappa \in \{4,9\}$

1.3.3 กำหนดค่าสัมประสิทธิ์ความเชื่อมั่น (Confidence Coefficient :$1-\alpha_0$) ที่ใช้ในการศึกษาเท่ากับ 0.95

1.3.4 ศึกษาประสิทธิภาพของตัวประมาณขนาดประชากรคือ $\hat{N}_{GT}, \hat{N}_{NewGT}, \hat{N}_{LB}$ และ $\hat{N}_{NewLB}$

1.3.5 ศึกษาประสิทธิภาพของช่วงเชื่อมั่นของขนาดประชากรที่สร้างจาก $\hat{N}_{GT}, \hat{N}_{NewGT}, \hat{N}_{LB}$ และ $\hat{N}_{NewLB}$

1.3.6 การศึกษาครั้งนี้จำลองสถานการณ์ให้ทำงานซ้ำ ๆ ทั้งหมด $B = 10,000$ รอบ

1.3.7 ประยุกต์ใช้ตัวประมาณขนาดประชากรกับข้อมูลจริง

1.4 เกณฑ์ที่ใช้ในการพิจารณา

เกณฑ์ที่ใช้ในการพิจารณาในงานวิจัยนี้ แบ่งออกเป็น 2 ส่วนดังนี้

ส่วนที่ 1 เกณฑ์ที่ใช้เปรียบเทียบประสิทธิภาพของตัวประมาณขนาดประชากร ภายใต้สถานการณ์ต่าง ๆ โดยพิจารณาความแม่น (Accuracy) จากค่าความเอนเอียงสัมพัทธ์ (Relative Bias)

$$RBias = \frac{\bar{\hat{N}} - N}{N}$$

โดยที่ $\bar{\hat{N}} = \frac{1}{B}\sum_{i=1}^{B}\hat{N}_i$ เมื่อ B แทนจำนวนรอบของการวนซ้ำๆ ในการจำลอง และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ (Relative Mean Square Error) ดังนี้

$$RMSE = \frac{1}{N^2}\sum_{i=1}^{B}\frac{(\hat{N}_i - N)^2}{B}$$

และพิจารณาความเที่ยง (Precision) จากค่าความแปรปรวนสัมพัทธ์ (Relative Variance) ดังนี้

$$RVar = \frac{1}{N^2}\sum_{i=1}^{B}\frac{(\hat{N}_i - \bar{\hat{N}})^2}{B}$$

Ref. code: 25595809030157UGW

โดยตัวประมาณขนาดประชากรที่มีประสิทธิภาพจะพิจารณาจากตัวประมาณขนาดประชากรที่ให้ค่าความเอนเอียงสัมพัทธ์ ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ และค่าความแปรปรวนสัมพัทธ์ต่ำสุด

ส่วนที่ 2 เกณฑ์ที่ใช้ในการเปรียบเทียบประสิทธิภาพของช่วงเชื่อมั่นของขนาดประชากรที่สร้างจากตัวประมาณขนาดประชากร

2.1 ค่าประมาณสัมประสิทธิ์ความเชื่อมั่น $(1-\hat{\alpha})$

พิจารณาค่าประมาณสัมประสิทธิ์ความเชื่อมั่น $(1-\hat{\alpha})$ โดยการหาช่วงเชื่อมั่น จากนั้นตรวจสอบช่วงเชื่อมั่นจากทั้งหมด 10,000 รอบ ถ้าช่วงเชื่อมั่นนั้นครอบคลุมค่าพารามิเตอร์จะให้ค่าเท่ากับ 1 ถ้าไม่ครอบคลุมค่าพารามิเตอร์ จะให้ค่าเป็น 0 และนับจำนวนครั้งบวกสะสมค่าไปเรื่อยๆ จนครบ 10,000 รอบของการทำซ้ำ นำจำนวนครั้งทั้งหมดที่ช่วงเชื่อมั่นครอบคลุมค่าพารามิเตอร์มาคำนวณหาค่าประมาณสัมประสิทธิ์ความเชื่อมั่น $(1-\hat{\alpha})$ จาก

$$1-\hat{\alpha}=\frac{C}{10,000}$$

โดยที่ C แทน จำนวนครั้งทั้งหมดที่ช่วงเชื่อมั่นครอบคลุมค่าพารามิเตอร์

การเปรียบเทียบค่าประมาณสัมประสิทธิ์ความเชื่อมั่นที่ได้กับเกณฑ์ที่กำหนดที่ได้จากการทดสอบสมมติฐาน พิจารณาจากการทดสอบสมมติฐาน โดยใช้ตัวสถิติทดสอบ Z (ในการทดสอบสมมติฐาน ผู้วิจัยได้กำหนดสัญลักษณ์ค่าประมาณสัมประสิทธิ์ความเชื่อมั่น เป็น $\hat{P}$) ดังนี้

$H_0 : P \geq P_0$ (ค่าสัมประสิทธิ์ความเชื่อมั่นไม่ต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด)

$H_1 : P < P_0$ (ค่าสัมประสิทธิ์ความเชื่อมั่นมีค่าต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด)

กำหนดระดับนัยสำคัญ $\alpha^* = 0.05$

ตัวสถิติที่ใช้ในการทดสอบ คือ

$$Z=\frac{\hat{P}-P_0}{\sqrt{\frac{P_0(1-P_0)}{M}}}$$

บริเวณวิกฤต คือ $Z < -Z_{\alpha^*}$

ดังนั้น จะปฏิเสธสมมติฐานหลัก H_0 เมื่อ $Z < -Z_{\alpha^*}$ หรือ $\frac{\hat{P}-P_0}{\sqrt{\frac{P_0(1-P_0)}{M}}} < -Z_{\alpha^*}$

Ref. code: 25595809030157UGW

นั่นคือ $\hat{P} < P_0 - Z_{\alpha^*}\sqrt{\frac{P_0(1-P_0)}{M}}$

ดังนั้น ตัวประมาณที่ให้ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นไม่ต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด คือตัวประมาณที่ให้ค่า

$$\hat{P} \geq P_0 - Z_{\alpha^*}\sqrt{\frac{P_0(1-P_0)}{M}}$$

เมื่อ P คือ สัมประสิทธิ์ความเชื่อมั่น หรือ $1-\alpha$

P_0 คือ สัมประสิทธิ์ความเชื่อมั่นที่กำหนด หรือ $1-\alpha_0$

$\hat{P}$ คือ ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นที่ได้จากการจำลอง หรือ $1-\hat{\alpha}$

M คือ จำนวนรอบของการทำซ้ำ (10,000 รอบ)

$\alpha*$ คือ ระดับนัยสำคัญในการทดสอบเท่ากับ 0.05

พิจารณาการสร้างเกณฑ์ที่ใช้ในการเปรียบเทียบค่าประมาณสัมประสิทธิ์ความเชื่อมั่นที่ระดับความเชื่อมั่น 95%

สมมติฐานการทดสอบ คือ

$H_0 : P \geq 0.95$

$H_1 : P < 0.95$

กำหนดระดับนัยสำคัญ $\alpha^* = 0.05$

นั่นคือ ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นไม่ต่ำกว่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด ถ้า

$$\hat{P} \geq 0.95 - 1.645\sqrt{\frac{0.95(1-0.95)}{10,000}}$$

$$\hat{P} \geq 0.9464$$

2.2 ค่าประมาณความกว้างเฉลี่ยของช่วงเชื่อมั่น (Average Length)

ค่าประมาณความกว้างเฉลี่ยของช่วงเชื่อมั่น คำนวณจากค่าเฉลี่ยของผลต่างระหว่างขีดจำกัดบนและขีดจำกัดล่างของช่วงเชื่อมั่นทั้งหมด 10,000 รอบ และนำผลต่างที่ได้แต่ละครั้งมาบวกกัน หารด้วยจำนวนครั้งที่ทำซ้ำ (10,000 รอบ)

$$AL = \frac{\sum_{i=1}^{10,000}(U_i - L_i)}{10,000} \qquad ; i = 1, 2, 3, ..., 10,000$$

โดยที่ U_i แทน ขีดจำกัดบนของช่วงเชื่อมั่นรอบที่ i

L_i แทน ขีดจำกัดล่างของช่วงเชื่อมั่นรอบที่ i

ดังนั้น ถ้าช่วงเชื่อมั่น 95 % ที่สร้างจากตัวประมาณขนาดประชากรใดที่ให้ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นไม่ต่ำกว่า 0.9464 จะถือว่าช่วงเชื่อมั่นที่สร้างจากตัวประมาณขนาดประชากรนั้นให้ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นไม่ต่ำกว่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด จากนั้นนำช่วงเชื่อมั่นที่สร้างจากตัวประมาณขนาดประชากรที่ให้ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นไม่ต่ำกว่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด ไปทำการเปรียบเทียบค่าประมาณความกว้างเฉลี่ยของช่วงเชื่อมั่น ช่วงเชื่อมั่นที่สร้างจากตัวประมาณขนาดประชากรใดให้ค่าประมาณความกว้างเฉลี่ยของช่วงเชื่อมั่นแคบที่สุด จะถือว่าช่วงเชื่อมั่นที่สร้างจากตัวประมาณขนาดประชากรนั้นเป็นช่วงเชื่อมั่นที่เหมาะสมสำหรับสถานการณ์ที่ศึกษา

1.5 ประโยชน์ที่ได้รับจากการศึกษา

1.5.1 ได้ตัวประมาณขนาดประชากรที่ปรับปรุงโดยใช้การถดถอยควอนไทล์ที่ระดับ 0.50

1.5.2 ได้ความแปรปรวนของตัวประมาณขนาดประชากรที่ปรับปรุงใหม่ที่ได้จากข้อ (1.5.1) เพื่อนำไปสร้างช่วงเชื่อมั่นของขนาดประชากร

1.5.3 เป็นแนวทางในการเลือกใช้ตัวประมาณขนาดประชากรที่เหมาะสมในสถานการณ์ต่างๆ

Ref. code: 25595809030157UGW

บทที่ 2
วรรณกรรมและงานวิจัยที่เกี่ยวข้อง

2.1 ตัวประมาณขนาดประชากรบนพื้นฐานการแจกแจงทวินามเชิงลบ

การแจกแจงปัวซง (Poisson distribution) ที่มีพารามิเตอร์ λ เมื่อพารามิเตอร์ λ มีค่าไม่คงที่หรือความน่าจะเป็นที่หน่วยตัวอย่างแต่ละหน่วยจะถูกพบ i ครั้งมีค่าไม่เท่ากัน ซึ่งขึ้นอยู่กับการแจกแจงของพารามิเตอร์ λ เรียกว่า Heterogeneous Poisson Model มีฟังก์ชันความน่าจะเป็นของตัวแปรสุ่มคือ

$$p_i = \int_0^\infty \frac{e^{-\lambda}\lambda^i}{i!} g(\lambda) d\lambda \qquad (2.1)$$

โดยที่ $g(\lambda)$ คือฟังก์ชันความหนาแน่นความน่าจะเป็น (Probability density function) ของตัวแปรสุ่ม Λ และเมื่อให้ Λ มีการแจกแจงแกมมา (Gamma distribution) ที่มีพารามิเตอร์ $\kappa, \theta = \frac{1-\pi}{\pi}$ ฟังก์ชันความหนาแน่นความน่าจะเป็นดังนี้

$$g(\lambda) = \frac{\theta^{-\kappa}\lambda^{\kappa-1}e^{\frac{-\lambda}{\theta}}}{\Gamma(\kappa)} \qquad (2.2)$$

โดยที่ $\lambda, \theta, \kappa > 0$

ความน่าจะเป็นที่หน่วยตัวอย่างแต่ละหน่วยจะถูกพบ i ครั้ง คือ

$$\begin{aligned} p_i &= \int_0^\infty \frac{e^{-\lambda}\lambda^i}{i!} g(\lambda) d\lambda \\ &= \int_0^\infty \frac{e^{-\lambda}\lambda^i}{i!} \cdot \frac{\theta^{-\kappa}\lambda^{\kappa-1}e^{\frac{-\lambda}{\theta}}}{\Gamma(\kappa)} d\lambda \\ &= \frac{\theta^{-\kappa}}{\Gamma(i+1)\Gamma(\kappa)} \int_0^\infty e^{-\lambda(1+\frac{1}{\theta})} \lambda^{i+\kappa-1} d\lambda \end{aligned}$$

กำหนดให้ $w = \lambda(1+\frac{1}{\theta}) = \lambda(\frac{\theta+1}{\theta})$ จะได้ $\lambda = \frac{w\theta}{\theta+1}$ และ $d\lambda = \frac{\theta}{\theta+1} dw$ ดังนั้น

Ref. code: 25595809030157UGW

$$= \frac{\theta^{-\kappa}}{\Gamma(i+1)\Gamma(\kappa)} \int_0^{\infty} e^{-w} (\frac{w\theta}{\theta+1})^{i+\kappa-1} \cdot (\frac{\theta}{\theta+1}) dw$$

$$= \frac{\theta^{-\kappa}}{\Gamma(i+1)\Gamma(\kappa)} \cdot (\frac{\theta}{\theta+1})^{i+\kappa} \int_0^{\infty} e^{-w} w^{(i+\kappa)-1} dw$$

เนื่องจาก $\int_0^{\infty} e^{-w} w^{(i+\kappa)-1} dw = \Gamma(i+\kappa)$ ดังนั้น

$$= \frac{\Gamma(i+\kappa)}{\Gamma(i+1)\Gamma(\kappa)} \theta^{-\kappa} (\frac{\theta}{\theta+1})^{i+\kappa}$$

$$= \frac{\Gamma(i+\kappa)}{\Gamma(i+1)\Gamma(\kappa)} (\frac{\theta}{\theta+1})^{i} (\frac{1}{\theta+1})^{\kappa}$$

จาก $\theta = \frac{1-\pi}{\pi}$ จะได้ $\theta + 1 = \frac{1}{\pi}$ และ $\frac{\theta}{\theta+1} = 1-\pi$ ดังนั้น

$$p_i = \frac{\Gamma(\kappa+i)}{\Gamma(i+1)\Gamma(\kappa)} \pi^{\kappa} (1-\pi)^i \qquad ,i = 0,1,2,...$$

$$\kappa > 0 \ , \ 0 \le \pi \le 1 \tag{2.3}$$

ซึ่งฟังก์ชันการแจกแจงนี้คือ การแจกแจงทวินามเชิงลบ (Negative Binomial distribution) ที่มีพารามิเตอร์ κ และ π

ตัวประมาณขนาดประชากรที่พัฒนาภายใต้การแจกแจงทวินามเชิงลบที่สนใจศึกษา ได้แก่ ตัวประมาณขนาดประชากร Generalized Turing และ ตัวประมาณขนาดประชากร Lanumteang&Böhning

2.1.1 ตัวประมาณขนาดประชากรของ Generalized Turing

ในปี ค.ศ. 2013 Böhning, Baksh, Lerdsuwansri และ Gallagher ได้เสนอตัวประมาณขนาดประชากร โดยกำหนดให้ Y มีการแจกแจงแบบปัวซง ที่มีพารามิเตอร์ λ โดยที่ $g(\lambda)$ มีการแจกแจงแกมมา ซึ่งจัดรูปได้ในรูปแบบของการแจกแจงทวินามเชิงลบ (Negative Binomial distribution) ที่มีพารามิเตอร์ κ และ π ดังสมการ (2.3)

$$p_i = \frac{\Gamma(\kappa+i)}{\Gamma(i+1)\Gamma(\kappa)} \pi^{\kappa} (1-\pi)^i \qquad ,i = 0,1,2,...$$

โดยที่ $\pi \in (0,1)$ และ $\kappa > 0$ และ $E(Y) = \frac{\kappa(1-\pi)}{\pi}$ และ $Var(Y) = \frac{\kappa(1-\pi)}{\pi^2}$

จากสมการ (2.3) สามารถเขียนให้อยู่ในรูปของฟังก์ชันหนาแน่นของอนุกรมกำลัง ได้ดังนี้

$$p_i = a_i t^i A(t) \tag{2.4}$$

โดย $a_i = \dfrac{\Gamma(\kappa + i)}{\Gamma(i+1)\Gamma(\kappa)}$ และ $A(t) = (1-t)^{\kappa}$

ดังนั้น $\dfrac{p_{i+1}}{p_i} = \dfrac{a_{i+1} t^{i+1} (1-t)^{\kappa}}{a_i t^i (1-t)^{\kappa}} = \dfrac{a_{i+1}}{a_i} t$ จัดรูปจะได้ $\dfrac{a_i p_{i+1}}{a_{i+1} p_i} = t = 1 - \pi$

พิจารณา

$$\begin{aligned} \frac{a_i}{a_{i+1}} &= \frac{\Gamma(\kappa+i)}{\Gamma(i+1)\Gamma(\kappa)} \cdot \frac{\Gamma(i+2)\Gamma(\kappa)}{\Gamma(i+1+\kappa)} \\ &= \frac{(i+\kappa-1)!(i+1)!}{i!(i+\kappa)!} \\ &= \frac{i+1}{i+\kappa} \end{aligned}$$

ดังนั้น

$$\frac{a_i p_{i+1}}{a_{i+1} p_i} = \left(\frac{i+1}{i+\kappa}\right) \frac{p_{i+1}}{p_i} = 1 - \pi$$

จัดรูปจะได้

$$(i+1)\frac{p_{i+1}}{p_i} = (1-\pi)(i+\kappa)$$

และเมื่อกำหนดให้ $r_i = (i+1)\dfrac{p_{i+1}}{p_i}$ จะได้

$$r_i = (i+1)\frac{p_{i+1}}{p_i} = (1-\pi)\kappa + (1-\pi)i \tag{2.5}$$

จากตัวแบบการถดถอยเชิงเส้นแสดงความสัมพันธ์ระหว่าง r_i กับ i คือ

$$r_i = \alpha + \beta i + \varepsilon_i \tag{2.6}$$

โดยที่ α, β คือ จุดตัดแกนและความชันตามลำดับ และ ε_i คือค่าคลาดเคลื่อน

จากสมการ (2.6) สามารถเขียนในรูปของการประมาณได้ดังนี้

$$\hat{r}_i = \hat{\alpha} + \hat{\beta} i \tag{2.7}$$

ซึ่งสามารถเขียนให้อยู่ในรูปของตัวแบบการถดถอยเชิงเส้นแสดงความสัมพันธ์ระหว่าง $\hat{r}_i$ กับ i ได้ดังนี้

$$\hat{r}_i = \underbrace{(1-\pi)\kappa}_{\hat{\alpha}} + \underbrace{(1-\pi)}_{\hat{\beta}} i \tag{2.8}$$

โดยที่ $\hat{\alpha} = (1-\pi)\kappa$ และ $\hat{\beta} = (1-\pi)$ และ $\hat{\alpha} > 0$ และ $0 < \hat{\beta} < 1$

ดังนั้นค่าประมาณของ κ มีค่า $\hat{\kappa} = \dfrac{\hat{\alpha}}{\hat{\beta}}$ โดยที่ $\hat{\alpha}$ และ $\hat{\beta}$ เป็นค่าประมาณสัมประสิทธิ์การถดถอยจากสมการถดถอยเชิงเส้นแบบถ่วงน้ำหนักด้วยเมทริกซ์ทแยงมุมที่ประกอบด้วยความแปรปรวนผกผันของ $\hat{r}_i$ $;w_i = \left[\hat{r}_i^2\left(\dfrac{1}{f_{i+1}} + \dfrac{1}{f_i}\right)\right]^{-1}$ (Böhning et al.,2008)

จากสมการ (2.3) จะได้ $p_0 = \pi^{\kappa}$, $p_1 = \kappa\pi^{\kappa}(1-\pi)$ และ $E(Y) = \dfrac{\kappa(1-\pi)}{\pi}$

ดังนั้น $\dfrac{p_1}{E(Y)} = \pi^{\kappa+1}$ และ $p_0 = \left[\pi^{\kappa+1}\right]^{\frac{\kappa}{\kappa+1}} = \left[\dfrac{p_1}{E(Y)}\right]^{\frac{\kappa}{\kappa+1}}$ จะได้ $\hat{p}_0 = \left(\dfrac{f_1/\hat{N}}{S/\hat{N}}\right)^{\frac{\kappa}{\kappa+1}} = \left(\dfrac{f_1}{S}\right)^{\frac{\kappa}{\kappa+1}}$ โดยที่ $S = \sum_{i=1}^{m} i f_i$ ได้ค่าประมาณของ p_0 คือ $\hat{p}_0 = (\dfrac{f_1}{S})^{\frac{\hat{\kappa}}{\hat{\kappa}+1}}$ จาก $\hat{N} = \dfrac{n}{1-\hat{p}_0}$ แทนค่า $\hat{p}_0$ จะได้ตัวประมาณขนาดประชากร Generalized Turing (Böhning et al.,2013) ดังนี้

$$\hat{N}_{GT} = \frac{n}{1-(\frac{f_1}{S})^{\frac{\hat{\kappa}}{\hat{\kappa}+1}}} \tag{2.9}$$

โดยที่ $S = \sum_{i=1}^{m} i f_i$ และ $\hat{\kappa} = \dfrac{\hat{\alpha}}{\hat{\beta}}$

ค่าประมาณของความแปรปรวนตัวประมาณขนาดประชากร Generalized Turing (Pijitrattana&Lerdsuwansri,2015) คือ

$$Var(\hat{N}_{GT}) = \frac{n\hat{p}_0}{(1-\hat{p}_0)^2} + \frac{n^2}{(1-\hat{p}_0)^4}\left(\frac{\hat{\kappa}}{\hat{\kappa}+1}\right)^2 f_1^{\frac{\hat{\kappa}-1}{\hat{\kappa}+1}} S^{-\frac{2\hat{\kappa}}{\hat{\kappa}+1}} \left\{\left(1-\frac{f_1}{\hat{N}}\right)+\hat{p}_0\right\} \quad (2.10)$$

โดยที่ $S=\sum_{i=1}^{m} i f_i$, $\hat{p}_0 = \left(\frac{f_1}{S}\right)^{\frac{\hat{\kappa}}{\hat{\kappa}+1}}$ และ $\hat{\kappa} = \frac{\hat{\alpha}}{\hat{\beta}}$

2.1.2 ตัวประมาณขนาดประชากร Lanumteang&Böhning

ในปี ค.ศ. 2011 Lanumteang และ Böhning ได้พัฒนาตัวประมาณขนาดประชากรภายใต้การแจกแจงปัวซงที่มีพารามิเตอร์ λ โดยกำหนดให้ λ มีการแจกแจงแกมมา ที่มีพารามิเตอร์ $\kappa, \theta = \frac{1-\pi}{\pi}$ ซึ่งทำให้ได้การแจกแจงทวินามเชิงลบ (Negative Binomial distribution) ที่มีพารามิเตอร์ κ และ π โดยใช้หน่วยตัวอย่างที่ถูกสุ่มพบครั้งที่ 1, 2 และ 3 และอาศัยลอการิทึมของอัตราส่วนของความถี่ที่เกิดขึ้นติดต่อกันและสมการถดถอยเชิงเส้นมาใช้พัฒนาตัวประมาณขนาดประชากร Lanumteang&Böhning

ฟังก์ชันการแจกแจงความน่าจะเป็นของการแจกแจงทวินามเชิงลบ จากสมการ (2.3)

$$p_i = \frac{\Gamma(\kappa+i)}{\Gamma(i+1)\Gamma(\kappa)} \pi^{\kappa} (1-\pi)^i \qquad ,i=0,1,2,...$$

กำหนดให้ $r_i = \frac{i p_i}{p_{i-1}}$ เมื่อแทนค่า p_i, p_{i-1} จะได้

$$r_i = (\kappa+i-1)(1-\pi) \quad (2.11)$$

จากสมการ (2.11) เขียนให้อยู่ในรูปสมการถดถอยเชิงเส้นของ $\log(r_i)$ กับ i และใช้อนุกรมเทเลอร์ของ $\log r_i$ รอบจุด $(\kappa-1)$ ได้ดังนี้

$$\log(r_i) = \log\left(k+i-1\right) + \log\left(1-\pi\right) \approx \underbrace{\log(1-\pi)+\log(\kappa-1)}_{\hat{\alpha}} + \underbrace{\frac{1}{\kappa-1}}_{\hat{\beta}} i \quad (2.12)$$

Ref. code: 25595809030157UGW

แทนค่า $i=2,3$ ในสมการ (2.12) เราจะได้ $\log r_2=\log(\frac{2f_2}{f_1})=\hat{\alpha}+2\hat{\beta}$ และ $\log r_3=\log(\frac{3f_3}{f_2})=\hat{\alpha}+3\hat{\beta}$ ซึ่งสามารถแก้สมการหาค่า $\hat{\alpha}$ และ $\hat{\beta}$ ได้ และแทนค่า $\hat{\alpha}$ และ $\hat{\beta}$ ในสมการ (2.12) จะได้ $\hat{\alpha}=3\log(\frac{2f_2}{f_1})-2\log(\frac{3f_3}{f_2})$ และ $\hat{\beta}=\log(\frac{3f_3}{f_2})-\log(\frac{2f_2}{f_1})$ ต่อมาแทน $i=1$ ในสมการ (2.12) จะได้

$$\log r_1=\log(\frac{f_1}{f_0})=\hat{\alpha}+\hat{\beta}$$

หรือเขียนได้อีกรูปแบบหนึ่ง ดังนี้

$$\log(\frac{f_1}{f_0})=\underbrace{3\log(\frac{2f_2}{f_1})-2\log(\frac{3f_3}{f_2})}_{\hat{\alpha}}+\underbrace{\log(\frac{3f_3}{f_2})-\log(\frac{2f_2}{f_1})}_{\hat{\beta}}=2\log(\frac{2f_2}{f_1})-\log(\frac{3f_3}{f_2})$$

จะได้

$$\log(f_0)=3\log f_1-2\log(2f_2)-\log(f_2)+\log(3f_3)=\log(\frac{3f_1^3f_3}{4f_2^3})$$

ดังนั้น ตัวประมาณขนาดประชากรของ Lanumteang&Böhning คือ

$$\hat{N}_{LB}=n+\frac{3f_1^3f_3}{4f_2^3} \qquad (2.13)$$

ค่าประมาณของความแปรปรวนตัวประมาณขนาดประชากรของ Lanumteang&Böhning คือ

$$Var(\hat{N}_{LB})=(\frac{9}{4})^2\frac{f_1^5f_3^2}{f_2^6}(\frac{f_1}{f_2}+1)+(\frac{3}{4})^2\frac{f_1^6f_3}{f_2^6}(1-\frac{f_3}{n})+(\frac{3n}{4})\frac{f_1^3f_3}{nf_2^3+\frac{3}{4}f_1^3f_3} \qquad (2.14)$$

2.2 คุณสมบัติ Monotonic non-decreasing

ในปี ค.ศ. 1987 Chao ได้พัฒนาตัวประมาณขนาดประชากรภายใต้ Heterogeneous Poisson Model โดยที่กำหนดให้ Y มีการแจกแจงปัวซงที่มีพารามิเตอร์ λ โดยที่ λ มีฟังก์ชันความหนาแน่นความน่าจะเป็นคือ $g(\lambda)$ ดังนั้น ความน่าจะเป็นที่หน่วยตัวอย่างแต่ละหน่วยจะถูกสุ่มพบ i ครั้ง คือ

Ref. code: 25595809030157UGW

$$p_i = \int_0^\infty \frac{e^{-\lambda}\lambda^i}{i!} g(\lambda)d\lambda \qquad ; i = 1, 2, 3, ...$$

โดยตัวประมาณขนาดประชากรของ Chao นั้นอาศัยอสมการ Cauchy-Schwartz ของตัวแปรสุ่ม X และ Y ได้ดังนี้

$$\left(E(XY)\right)^2 \le E\left(X^2\right)E\left(Y^2\right) \tag{2.15}$$

กำหนดให้ $X = \mu(\lambda)$, $Y = \nu(\lambda)$ คือฟังก์ชันของ λ และสมมติให้ $g(\lambda)$ อยู่ในช่วง $(0,\infty)$ จะได้

$$\left(\int_0^\infty \mu(\lambda)\nu(\lambda)g(\lambda)d\lambda\right)^2 \le \left(\int_0^\infty \mu(\lambda)^2 g(\lambda)d\lambda\right)\left(\int_0^\infty \nu(\lambda)^2 g(\lambda)d\lambda\right)$$

และให้ $\mu(\lambda) = (e^{-\lambda}\lambda^{i-1})^{\frac{1}{2}}$ และ $\nu(\lambda) = (e^{-\lambda}\lambda^{i+1})^{\frac{1}{2}}$ และ $\mu(\lambda)\nu(\lambda) = e^{-\lambda}\lambda^i$ แทนค่า ได้ดังนี้

$$\left(\int_0^\infty e^{-\lambda}\lambda^i g(\lambda)d\lambda\right)^2 \le \left(\int_0^\infty e^{-\lambda}\lambda^{i-1} g(\lambda)d\lambda\right)\left(\int_0^\infty e^{-\lambda}\lambda^{i+1} g(\lambda)d\lambda\right)$$

$$\left(i!\int_0^\infty \frac{e^{-\lambda}\lambda^i}{i!} g(\lambda)d\lambda\right)^2 \le \left((i-1)!\int_0^\infty \frac{e^{-\lambda}\lambda^{i-1}}{(i-1)!} g(\lambda)d\lambda\right)\left((i+1)!\int_0^\infty \frac{e^{-\lambda}\lambda^{i+1}}{(i+1)!} g(\lambda)d\lambda\right)$$

$$(i!\,p_i)^2 \le (i-1)!\,p_{i-1}(i+1)!\,p_{i+1}$$

$$\frac{ip_i}{p_{i-1}} \le \frac{(i+1)p_{i+1}}{p_i} \tag{2.16}$$

แทนค่า $i = 1,2,3,...$ ในสมการ (2.16) จะได้ $\frac{p_1}{p_0} \le \frac{2p_2}{p_1} \le \frac{3p_3}{p_2} \le \frac{4p_4}{p_4} \le ...$ ซึ่งเป็นคุณสมบัติ Monotonic non-decreasing ภายใต้การแจกแจงแบบ Heterogeneous Poisson Model และเมื่อ $r_i = (i+1)\frac{p_{i+1}}{p_i}$ เมื่อเทียบกับสมการ (2.16) สามารถสรุปได้ว่า $r_i = (i+1)\frac{p_{i+1}}{p_i}$ นั้นมีคุณสมบัติ Monotonic non-decreasing

2.3 Ratio Plot และตัวแบบการถดถอยเชิงเส้น

การใช้ Ratio Plot และตัวแบบการถดถอยเชิงเส้นที่อยู่บนพื้นฐานของการแจกแจงทวินามเชิงลบ เพื่อใช้ในการประมาณขนาดประชากร เริ่มต้นจาก Chao (Chao,1987) ที่ได้พัฒนาตัวประมาณขนาดประชากรภายใต้ Heterogeneous Poisson Model โดยอาศัยอสมการ Cauchy-Schwartz ซึ่งพบว่าจากการใช้อสมการ Cauchy-Schwartz นั้นมีคุณสมบัติ Monotonic non-decreasing ซึ่งพิจารณาได้จาก

$$\frac{p_1}{p_0} \le \frac{2p_2}{p_1} \le \frac{3p_3}{p_2} \le \frac{4p_4}{p_4} \le \ldots$$

ต่อมาในปี ค.ศ. 2013 Böhning และคณะ กำหนดให้ $r_i = \frac{(i+1)p_{i+1}}{p_i}$ โดยค่าประมาณของ r_i หาได้จาก $\hat{r}_i = \frac{(i+1)f_{i+1}}{f_i}$ และนำเสนอเครื่องมือที่เรียกว่า Ratio Plot ซึ่งเป็นการพล๊อตกราฟระหว่าง $\hat{r}_i$ กับ i เพื่อใช้ในการตรวจสอบว่าข้อมูลแจงนับมีลักษณะเป็น Homogeneous Poisson Model หรือ Heterogeneous Poisson Model ในกรณีที่ข้อมูลแจงนับมีลักษณะเป็น Homogeneous Poisson Model กราฟ Ratio Plot จะแสดงเส้นตรงที่มีความชันเป็นศูนย์ และเมื่อข้อมูลแจงนับมีลักษณะเป็น Heterogeneous Poisson Model กราฟจะแสดงคุณสมบัติ Monotonic non-decreasing นอกจากนี้ Böhning และคณะได้นำเสนอ Ratio Plot ภายใต้การแจกแจงทวินามเชิงลบ (Negative Binomial distribution) ที่มีพารามิเตอร์ κ และ π ซึ่งมีฟังก์ชันการแจกแจงความน่าจะเป็นดังนี้

$$p_i = \frac{\Gamma(\kappa+i)}{\Gamma(i+1)\Gamma(\kappa)}\pi^{\kappa}(1-\pi)^i \qquad ,i = 0,1,2,\ldots$$

จะได้

$$r_i = \frac{(i+1)p_{i+1}}{p_i} = \left(1-\pi\right)\kappa + (1-\pi)\kappa$$

จากสมการข้างต้น จะเห็นว่าเป็นสมการเส้นตรงแสดงความสัมพันธ์ระหว่าง r_i กับ i ที่มีจุดตัดแกน $(1-\pi)\kappa$ และความชัน $(1-\pi)$ และเมื่อพล๊อตกราฟ Ratio Plot จะแสดงเส้นตรงที่มีความชันเป็นบวก หรือเรียกว่าข้อมูลแจงนับที่อยู่บนพื้นฐานของการแจกแจงทวินามเชิงลบนั้น มีลักษณะ Heterogeneous Poisson Model

2.4 การถดถอยควอนไทล์

ในปี ค.ศ. 1978 Koenker and Bassett ได้นำเสนอการวิเคราะห์การถดถอยที่เรียกว่าการถดถอยควอนไทล์ (Quantile Regression) ซึ่งเป็นวิธีการทางสถิติที่ใช้ในการศึกษาความสัมพันธ์ของตัวแปรตั้งแต่สองตัวขึ้นไปและประมาณค่าพารามิเตอร์ในตัวแบบ

การถดถอยควอนไทล์ (Quantile Regression) เป็นการศึกษาความสัมพันธ์ของตัวแปรเช่นเดียวกับการวิเคราะห์การถดถอยเชิงเส้นอย่างง่าย (Simple Linear Regression) ที่มีพื้นฐานการประมาณค่าพารามิเตอร์ในตัวแบบจากการถดถอยเข้าสู่ค่าเฉลี่ยของตัวแปร Y แต่การวิเคราะห์การถดถอยควอนไทล์มีพื้นฐานมาจากการถดถอยเข้าสู่ค่ามัธยฐานหรือขึ้นอยู่กับระดับของควอนไทล์ที่กำหนด โดยระดับของควอนไทล์ พิจารณาจากฟังก์ชันการแจกแจงความน่าจะเป็นของตัวแปรสุ่ม Y คือ $F(y)$

$$F(y) = \Pr ob(Y \leq y)$$

โดยที่ Y คือ ตัวแปรสุ่ม

y คือ ระดับที่ต้องการหาความน่าจะเป็น

ดังนั้น ควอนไทล์ที่ τ ของตัวแปร Y ซึ่งเป็นฟังก์ชันผกผัน (inverse function) แสดงได้ดังนี้

$$Q(\tau) = \inf\{y : F(y) \geq \tau\}$$

โดยที่ τ คือ ระดับควอนไทล์ โดยที่ $\tau \in (0,1)$

$Q(\tau)$ คือ ฟังก์ชันผกผันที่ควอนไทล์ที่ τ

และตัวแบบของการถดถอยควอนไทล์ แสดงได้ดังนี้

$$y_\tau = \alpha + \beta x_\tau + \varepsilon_\tau \qquad (2.17)$$

หรือเขียนได้รูปแบบหนึ่ง ได้ดังนี้

$$y_\tau = q_\tau(Y \mid X) + \varepsilon_\tau \qquad (2.18)$$

โดยที่ y_τ คือ ค่าของตัวแปรตามระดับที่ τ

α, β คือ พารามิเตอร์ที่เรียกว่าสัมประสิทธิ์การถดถอย (regression coefficient)

x_τ คือ ตัวแปรอิสระระดับที่ τ

Ref. code: 25595809030157UGW

$q_\tau(Y|X)$ คือ ค่าควอนไทล์ระดับที่ τ

ε_τ คือ ค่าคลาดเคลื่อนระดับที่ τ โดยที่ $\tau \in (0,1)$

การประมาณค่าพารามิเตอร์ α, β ของการถดถอยแบบควอนไทล์อาศัยวิธี Least Absolute Deviation (LAD) เป็นวิธีที่ใช้ในการประมาณค่าพารามิเตอร์ที่อยู่ในสมการถดถอย โดยเป็นวิธีที่ทำให้ผลรวมค่าสัมบูรณ์ของส่วนเบี่ยงเบนมีค่าน้อยที่สุด นั่นคือ

$$\hat{\beta} = \arg\min \sum_{i=1}^{n} |y_i - \hat{y}_i| \tag{2.19}$$

เนื่องจากวิธี Least Absolute Deviation นั้นไม่สามารถหาตัวประมาณพารามิเตอร์ในตัวแบบในรูปแบบปิด (Closed Form) ได้ ดังนั้น ในที่นี้เราจึงอาศัยการประมาณค่าพารามิเตอร์ในตัวแบบจากโปรแกรมสำเร็จรูป R Studio โดยที่ Koenker and Bassett(1978) ได้สร้างอัลกอริทึมที่สามารถหาตัวประมาณพารามิเตอร์ในตัวแบบไว้ 3 อัลกอริทึม ได้แก่ Simplex method , Interior method และ Smoothing algorithm อย่างไรก็ดี ในการประมาณพารามิเตอร์ในตัวแบบโดยใช้อัลกอริทึมที่แตกต่างกัน ค่าประมาณที่ได้มีค่าแตกต่างกันเล็กน้อย เนื่องจากอัลกอริทึม Simplex method เป็นอัลกอริทึมที่ตั้งค่าเป็นอัลกอริทึมพื้นฐานในโปรแกรม R Studio และถูกนำมาใช้มากที่สุดในงานวิจัยอื่นๆ ด้วยเหตุนี้ เราจึงใช้อัลกอริทึม Simplex method ในการประมาณค่าพารามิเตอร์ในตัวแบบของการถดถอยควอนไทล์ในงานวิจัยนี้

2.5 การหาความแปรปรวนของตัวประมาณขนาดประชากร

ในปี ค.ศ. 2008 Böhning ได้นำเสนอการหาความแปรปรวนของตัวประมาณขนาดประชากรของวิธี Capture-Recapture โดยใช้ค่าคาดหมายแบบมีเงื่อนไข (Conditioning Expectation) ซึ่งในการหาความแปรปรวนของตัวประมาณขนาดประชากรโดยใช้ค่าคาดหมายแบบมีเงื่อนไขนั้น จะอาศัยแหล่งที่มาจาก 2 แหล่ง ดังนี้

1. ความแปรปรวนเนื่องจากการประมาณค่าพารามิเตอร์ในตัวแบบ

และ 2. ความแปรปรวนเนื่องจากขนาดของจำนวนหน่วยตัวอย่าง (n)

โดยที่ $\hat{N}$ คือ ตัวประมาณขนาดประชากรที่สนใจศึกษา

ดังนั้น ความแปรปรวนของตัวประมาณขนาดประชากร แสดงได้ดังนี้

Ref. code: 25595809030157UGW

$$Var(\hat{N})=Var_n\left\{E(\hat{N}\,|\,n)\right\}+E_n\left\{Var(\hat{N}\,|\,n)\right\} \tag{2.20}$$

โดยที่ E_n และ Var_n คือค่าเฉลี่ยและความแปรปรวนที่ขึ้นอยู่กับการแจกแจงของ n โดยที่ n มีการแจกแจงทวินาม (Binomial distribution) ที่มีพารามิเตอร์ N และ $1-p_0$

2.6 วิธีเดลต้า (Delta method)

วิธีเดลต้า (Delta method) เป็นวิธีที่ใช้หาตัวประมาณค่าเฉลี่ยและตัวประมาณความแปรปรวนของฟังก์ชันตัวแปรสุ่ม โดยอาศัยการกระจายอนุกรมเทย์เลอร์ (Taylor Expansion)

กำหนดให้ X เป็นตัวแปรสุ่มที่กำหนดให้ค่าเฉลี่ย $(E(X))$ และความแปรปรวน $(Var(X))$ นั้นทราบค่า และตัวแปรสุ่ม Y เป็นฟังก์ชันที่อยู่ในรูปของตัวแปรสุ่ม X โดยกำหนดให้ $E(X)=\mu_X$ เมื่อกระจายด้วยอนุกรมเทย์เลอร์อันดับที่ 1 รอบค่าเฉลี่ยจะได้

$$\begin{aligned} Y &= g(X) \\ &\approx g(u_X)+(X-\mu_X)g'(\mu_X) \end{aligned}$$

ดังนั้น $E(Y)$ และ $Var(Y)$ หาได้จาก

$$E(Y)\approx E[g(u_X)]+E[(X-\mu_X)g'(\mu_X)]=g(\mu_X)$$

$$Var(Y)\approx Var[g(u_X)]+Var[(X-\mu_X)g'(\mu_X)]=Var(X)\left[g'(\mu_X)\right]^2$$

เนื่องจาก $E(g(X))\neq g(E(X))$ ดังนั้นต้องอาศัยการกระจายด้วยอนุกรมเทย์เลอร์อันดับที่ 2 รอบค่าเฉลี่ยเพื่อหา $E(Y)$

จาก $\quad Y=g(X)$

$$Y\approx g(u_X)+(X-\mu_X)g'(\mu_X)+\frac{1}{2}(X-\mu_X)^2 g''(\mu_X)$$

$$E(Y)\approx E[g(u_X)]+E[(X-\mu_X)g'(\mu_X)]+\frac{1}{2}E[(X-\mu_X)^2 g''(\mu_X)]$$

$$=g(u_X)+\frac{1}{2}Var(X)g''(\mu_X)$$

ดังนั้นหา $E(Y)$ และ $Var(Y)$ จากวิธีเดลต้าได้ดังนี้

$$E(Y) \approx \left[g(\mu_X)\right] + \frac{1}{2}Var(X)g''(\mu_X) \tag{2.21}$$

$$Var(Y) \approx \left[g'(\mu_X)\right]^2 Var(X) \tag{2.22}$$

2.7 งานวิจัยที่เกี่ยวข้อง

ในปี ค.ศ. 1987 Chao ได้เสนอตัวประมาณขนาดประชากรภายใต้สมมติฐานที่ว่า Y มีการแจกแจงปัวซงที่มีลักษณะ Heterogeneous Poisson Model ขึ้น โดยเลือกใช้จำนวนหน่วยตัวอย่างที่ถูกสุ่มพบ 1 และ 2 ครั้ง โดยอาศัยอสมการของ Cauchy-Schwartz ของตัวแปรสุ่ม X และ Y ในการหาตัวประมาณขนาดประชากร และใช้ข้อมูลเฉพาะจำนวนหน่วยตัวอย่างที่ถูกพบ 1 และ 2 ครั้งเท่านั้น ซึ่งตัวประมาณขนาดประชากรของ Chao คือ $\hat{N}_{Chao} = n + \frac{f_1^2}{2f_2}$ จากตัวประมาณขนาดประชากรของ Chao พบว่าจากการใช้อสมการ Cauchy-Schwartz นั้นมีคุณสมบัติ Monotonic non-decreasing ซึ่งพิจารณาได้จาก $\frac{p_1}{p_0} \le \frac{2p_2}{p_1} \le \frac{3p_3}{p_2} \le \frac{4p_4}{p_4} \le \ldots$ ซึ่งคุณสมบัติ Monotonic non-decreasing ที่ได้จากการใช้อสมการ Cauchy-Schwartz ของตัวประมาณขนาดประชากรของ Chao ใช้พิสูจน์ว่าข้อมูลแจงนับที่สนใจศึกษามีลักษณะ Heterogeneous Poisson Model

ในปี ค.ศ. 2008 Böhning ได้เสนอการหาค่าความแปรปรวนของตัวประมาณขนาดประชากรของวิธี Capture-Recapture โดยใช้ค่าคาดหมายแบบมีเงื่อนไข (Conditioning Expectation) ซึ่งในการหาความแปรปรวนของตัวประมาณขนาดประชากรโดยใช้ค่าคาดหมายแบบมีเงื่อนไขนั้น อาศัยแหล่งที่มาจาก 2 แหล่ง ได้แก่ ความแปรปรวนเนื่องจากการประมาณค่าพารามิเตอร์ในตัวแบบ และความแปรปรวนเนื่องจากขนาดของจำนวนหน่วยตัวอย่าง ซึ่งปัจจุบัน ค่าความแปรปรวนขนาดประชากรโดยใช้ค่าคาดหมายแบบมีเงื่อนไขถูกนำมาใช้อย่างแพร่หลาย

ในปี ค.ศ. 2011 Lanumteang&Böhning ได้พัฒนาตัวประมาณขนาดประชากรภายใต้ การแจกแจงปัวซงที่มีพารามิเตอร์ λ โดยกำหนดให้ λ มีการแจกแจงแกมมา ที่มีพารามิเตอร์ κ และ $\theta = \frac{1-\pi}{\pi}$ ซึ่งทำให้ได้การแจกแจงทวินามเชิงลบ (Negative Binomial distribution) ที่มีพารามิเตอร์ κ และ π โดยใช้หน่วยตัวอย่างที่ถูกสุ่มพบครั้งที่ 1, 2 และ 3 และอาศัยลอการิทึมของอัตราส่วนของความถี่ที่เกิดขึ้นติดต่อกันและสมการถดถอยเชิงเส้นมาใช้พัฒนาตัวประมาณขนาดประชากร Lanumteang&Böhning และได้นำตัวประมาณขนาดประชากรที่ได้พัฒนามาประยุกต์ใช้กับข้อมูล

Ref. code: 25595809030157UGW

จริง ผลลัพธ์ที่ได้พบว่า ตัวประมาณขนาดประชากร Lanumteang&Böhning มีค่าความเอนเอียงต่ำ แต่พบว่าความแปรปรวนของตัวประมาณขนาดประชากร Lanumteang&Böhning ยังมีค่าสูงกว่าตัวประมาณอื่นๆ ในทุกสถานการณ์

ในปี ค.ศ. 2013 Böhning, Baksh, Lerdsuwansri และ Gallagher ได้นำเสนอเครื่องมือที่เรียกว่า Ratio Plot เพื่อใช้ตรวจสอบข้อมูลแจงนับว่ามีลักษณะ Homogeneous Poisson Model หรือ Heterogeneous Poisson Model โดยการพล็อตกราฟความสัมพันธ์ระหว่าง $\hat{r}_x = \frac{(x+1)f_{x+1}}{f_x}$ กับ x ในกรณีที่ข้อมูลแจงนับมีลักษณะ Homogeneous Poisson Model กราฟ Ratio Plot แสดงเส้นตรงที่มีความชันเป็นศูนย์ และกรณีที่ข้อมูลแจงนับมีลักษณะ Heterogeneous Poisson Model กราฟ Ratio Plot จะแสดงคุณสมบัติ monotonic non-decreasing หรือแสดงเส้นตรงที่มีความชันเป็นบวก นอกจากนี้ Böhning และคณะได้พัฒนาตัวประมาณขนาดประชากร Generalized Turing และนำเสนอกราฟ Ratio Plot ภายใต้การแจกแจงทวินามเชิงลบที่มีพารามิเตอร์ κ และ π จาก $p_x = \frac{\Gamma(\kappa+x)}{\Gamma(x+1)\Gamma(\kappa)}\pi^{\kappa}(1-\pi)^x$ และจาก $r_x = \frac{(x+1)p_{x+1}}{p_x}$ จัดรูปจะได้ $r_x = (1-\pi)\kappa + (1-\pi)x$ ซึ่งเขียนอยู่ในรูปของสมการเส้นตรงแสดงความสัมพันธ์ระหว่าง r_x กับ x ซึ่งกราฟแสดงความชันเป็นบวก นั่นคือลักษณะ Heterogeneous Poisson Model

ในปี ค.ศ. 2014 Pijitrattana และ Lerdsuwansri ได้ต่อยอดแนวคิดจาก Böhning และคณะ (2013) ที่ได้นำเสนอตัวประมาณขนาดประชากร Generalized Turing แต่ยังไม่มีการศึกษาการประมาณค่าแบบช่วง ดังนั้น Pijitrattana และ Lerdsuwansri จึงได้เสนอการสร้างช่วงเชื่อมั่นของขนาดประชากรโดย Generalized Turing โดยวิธีของวาลด์ (Wald Method) และอาศัยการหาค่าความแปรปรวนของตัวประมาณขนาดประชากรโดยวิธีค่าคาดหมายแบบมีเงื่อนไข (conditioning technique) และเปรียบเทียบประสิทธิภาพของช่วงเชื่อมั่นของขนาดประชากรโดย Generalized Turing, Chao, Zelterman และ Lanumteang&Böhning พบว่าช่วงเชื่อมั่นของขนาดประชากรโดย Generalized Turing มีประสิทธิภาพดี เมื่อประชากรอยู่ภายใต้การแจกแจงทวินามเชิงลบที่มีค่าเฉลี่ยมีค่ามาก

ในปี ค.ศ. 1978 Koenker และ Bassett ได้นำเสนอการถดถอยควอนไทล์ขึ้น ซึ่งเป็นวิธีทางสถิติที่ใช้ในศึกษาความสัมพันธ์ของตัวแปร 2 ตัวขึ้นไป โดยขึ้นอยู่กับการกำหนดระดับของควอนไทล์ Koenker และ Bassett ได้นำการถดถอยควอนไทล์มาใช้ในการศึกษาความสัมพันธ์ของตัวแปร แต่ละระดับของควอนไทล์ และได้อธิบายที่มาของการประมาณค่าพารามิเตอร์ในตัวแบบของการถดถอยควอนไทล์

ในปี ค.ศ. 2012 กัญญ์สุดา และคณะได้ศึกษาผลกระทบของหนี้สาธารณะที่มีผลต่ออัตราการเจริญเติบโตทางเศรษฐกิจของกลุ่มประเทศต่างๆ ด้วยวิธีการถดถอยแบบพาแนลควอนไทล์ โดยทำ

การเปรียบเทียบผลกับการวิธีการถดถอยโดยทั่วไป (condition mean) และเปรียบเทียบผลกระทบดังกล่าวของกลุ่มประเทศต่างๆ ที่มีระดับหนี้สาธารณะที่แตกต่างกันของ 156 ประเทศ โดยใช้ข้อมูลรายปีตั้งแต่ปี 1981–2010 ผลการศึกษา พบว่า หนี้สาธารณะมีผลกระทบทางลบต่ออัตราการเจริญเติบโตทางเศรษฐกิจในทุกกลุ่มประเทศ แต่บางกลุ่มประเทศอาจมีผลกระทบแค่บางช่วงของอัตราการเจริญเติบโตทางเศรษฐกิจเท่านั้น และกลุ่มประเทศเกือบทุกกลุ่มมีผลการศึกษาสอดคล้องกับวิธีการถดถอยโดยทั่วไป

ในปี ค.ศ. 2015 ศราวุฒิ ชุมภูอินตา ได้ศึกษาปัจจัยทางเศรษฐกิจมหภาคที่มีผลต่ออัตราผลตอบแทนของดัชนีราคาหุ้นกลุ่มพลังงานโดยวิธีการถดถอยแบบควอนไทล์ ที่ระดับควอนไทล์ 0.25, 0.50 และ 0.75 ซึ่งนำไปเปรียบเทียบผลกระทบของแต่ละปัจจัยเศรษฐกิจมหภาคที่มีผลต่ออัตราผลตอบแทนของดัชนีราคาหุ้นกลุ่ม พลังงานตามระดับควอนไทล์ที่แตกต่างกันและได้ใช้ ข้อมูลทุติยภูมิรายไตรมาสตั้งแต่ไตรมาสที่ 1 ปี 2547 ถึงไตรมาสที่ 4 ปี 2557 โดยใช้ อัตราดอกเบี้ย, อัตราการเปลี่ยนแปลงของอัตราแลกเปลี่ยนเงินตราต่างประเทศ, อัตราเงินเฟ้อ และอัตราการเปลี่ยนแปลงราคาน้ำมันดิบในตลาดดูไบ เป็นตัวแปรอิสระ และใช้อัตราผลตอบแทนของดัชนีราคาหุ้นกลุ่มพลังงานเป็นตัวแปรตาม จากการทดสอบพบว่ามีเพียงอัตราการเปลี่ยนแปลงราคาน้ำมันดิบในตลาดดูไบที่มีผลต่ออัตราผลตอบแทนของดัชนีราคาหุ้นกลุ่มพลังงานในทิศเดียวกันที่ระดับควอนไทล์ 0.25 , 0.50 และ 0.75 โดยมีผลกระทบมากขึ้นเมื่ออัตราผลตอบแทนของดัชนีราคาหุ้นกลุ่มพลังงานเพิ่มขึ้น และมีอัตราดอกเบี้ยที่ส่งผลต่ออัตราผลตอบแทนของดัชนีราคาหุ้นกลุ่มพลังงานในทิศตรงข้ามที่ระดับควอนไทล์ 0.50 เท่านั้น ส่วนตัวแปรอื่นไม่มีนัยสำคัญที่ระดับนัยสำคัญ

บทที่ 3
วิธีการวิจัย

งานวิจัยนี้มีวัตถุประสงค์เพื่อปรับปรุงตัวประมาณขนาดประชากรที่อยู่บนพื้นฐานของการแจกแจงทวินามเชิงลบ (Negative Binomial distribution) โดยใช้การถดถอยควอนไทล์ที่ระดับ 0.50 ซึ่งตัวประมาณขนาดประชากรที่สนใจศึกษาคือ ตัวประมาณขนาดประชากร Generalized Turing ($\hat{N}_{GT}$) และตัวประมาณขนาดประชากร Lanumteang&Böhning ($\hat{N}_{LB}$) พร้อมทั้งหาความแปรปรวนของตัวประมาณขนาดประชากรที่ได้พัฒนาขึ้นใหม่เพื่อนำมาใช้ประโยชน์ในการสร้างช่วงเชื่อมั่นของขนาดประชากรต่อไป

3.1 ตัวประมาณขนาดประชากร Generalized Turing ที่ปรับปรุงโดยใช้การถดถอยควอนไทล์ที่ระดับ 0.50

จากตัวแบบการถดถอยเชิงเส้นแสดงความสัมพันธ์ระหว่าง $\hat{r}_i$ กับ i สมการ (2.7)

$$\hat{r}_i = \hat{\alpha} + \hat{\beta} i$$

ซึ่งจัดรูปได้ดังสมการ (2.8)

$$\hat{r}_i = \underbrace{(1-\pi)\kappa}_{\hat{\alpha}} + \underbrace{(1-\pi)}_{\hat{\beta}} i$$

โดยค่าประมาณ $\hat{\kappa}$ ของตัวประมาณขนาดประชากร Generalized Turing หาได้จาก $\hat{\kappa} = \dfrac{\hat{\alpha}}{\hat{\beta}}$ โดยที่ $\hat{\alpha}$ และ $\hat{\beta}$ เป็นค่าประมาณสัมประสิทธิ์การถดถอยจากสมการถดถอยเชิงเส้นแบบถ่วงน้ำหนักด้วยเมทริกซ์ทแยงมุมที่ประกอบด้วยความแปรปรวนผกผันของ $\hat{r}_i$ $;w_i = \left[\hat{r}_i^2\left(\dfrac{1}{f_{i+1}} + \dfrac{1}{f_i}\right)\right]^{-1}$ (Böhning et al.,2008) และจากสูตรการหาตัวประมาณขนาดประชากร Generalized Turing คือ

$$\hat{N}_{GT} = \frac{n}{1-(\frac{f_1}{S})^{\frac{\hat{\kappa}}{\hat{\kappa}+1}}}$$

เนื่องจากผู้วิจัยต้องการปรับปรุงตัวประมาณขนาดประชากร Generalized Turing โดยประมาณค่า $\hat{\alpha}$ และ $\hat{\beta}$ จากการถดถอยควอนไทล์ที่ระดับ 0.50

ดังนั้น ตัวประมาณขนาดประชากร Generalized Turing ที่ปรับปรุงโดยใช้การถดถอย

Ref. code: 25595809030157UGW

ควอนไทล์ที่ระดับ 0.50 คือ

$$\hat{N}_{NewGT} = \frac{n}{1-(\frac{f_1}{S})^{\frac{\hat{\kappa}}{\hat{\kappa}+1}}} \tag{3.1}$$

โดยที่ $\hat{\kappa} = \frac{\hat{\alpha}}{\hat{\beta}}$ และ $\hat{\alpha}, \hat{\beta}$ เป็นค่าประมาณสัมประสิทธิ์การถดถอยที่ได้จากการถดถอยควอนไทล์ที่ระดับ 0.50 โดยประมาณด้วยอัลกอริทึม Simplex method

3.2 ตัวประมาณขนาดประชากร Lanumteang&Böhning ที่ปรับปรุงโดยใช้การถดถอยควอนไทล์ที่ระดับ 0.50

จากสมการ (2.3)

$$p_i = \frac{\Gamma(\kappa+i)}{\Gamma(i+1)\Gamma(\kappa)}\pi^{\kappa}(1-\pi)^i \qquad ,i = 0,1,2,...$$

โดยที่ $\kappa > 0$ และ $\pi \in (0,1)$

และจากสมการ (2.11)

$$r_i = \frac{ip_i}{p_{i-1}} = (\kappa+i-1)(1-\pi)$$

และแปลงให้อยู่ในรูปของ $\log(r_i)$ กับ i และใช้อนุกรมเทเลอร์ของ $\log(r_i)$ รอบจุด $(\kappa-1)$ ได้ดังนี้

$$\log(r_i) = \log(\kappa+i-1) + \log(1-\pi) \approx \underbrace{\log(1-\pi)+\log(\kappa-1)}_{\alpha} + \underbrace{\frac{1}{\kappa-1}}_{\beta} i$$

โดยที่ $\alpha = \log(1-\pi)+\log(\kappa-1)$, $\beta = \frac{1}{\kappa-1}$ และ $\alpha, \beta > 0$

จากสมการข้างต้น สามารถเขียนในรูปของตัวแบบถดถอยเชิงเส้นแสดงความสัมพันธ์ระหว่าง $\log(\hat{r}_i)$ กับ i ได้ดังนี้

$$\log(\hat{r}_i) = \hat{\alpha} + \hat{\beta} i \tag{3.2}$$

โดยที่ $\hat{\alpha},\hat{\beta}$ คือ จุดตัดแกนและความชัน ตามลำดับ และ $\hat{\alpha},\hat{\beta}$ เป็นค่าประมาณสัมประสิทธิ์การถดถอยที่ได้จากการถดถอยควอนไทล์ที่ระดับ 0.50
และสมการถดถอยเชิงเส้นแสดงความสัมพันธ์ระหว่าง $\log(\hat{r}_i)$ กับ i สามารถเขียนในรูปการประมาณได้ดังนี้

$$\log(\hat{r}_i) = \log(\frac{if_i}{f_{i-1}}) \quad (3.3)$$

แทนค่า $i=1$ ในสมการ (3.2) และ (3.3) จะได้

$$\log(\frac{f_1}{f_0}) = \hat{\alpha} + \hat{\beta}$$

ดังนั้นจำนวนหน่วยตัวอย่างที่ไม่ถูกพบ หรือ f_0 หาได้จาก

$$\frac{f_1}{f_0} = e^{\hat{\alpha}+\hat{\beta}}$$

จะได้

$$\hat{f}_0 = \frac{f_1}{e^{\hat{\alpha}+\hat{\beta}}}$$

ดังนั้น ตัวประมาณขนาดประชากร Lanumteang&Böhing ที่ปรับปรุงโดยใช้การถดถอยควอนไทล์ที่ระดับ 0.50 คือ

$$\hat{N}_{NewLB} = n + \frac{f_1}{e^{\hat{\alpha}+\hat{\beta}}} \quad (3.4)$$

โดยที่ $\hat{\alpha}$ และ $\hat{\beta}$ เป็นค่าประมาณสัมประสิทธิ์การถดถอยที่ได้จากการถดถอยควอนไทล์ที่ระดับ 0.50 โดยประมาณด้วยอัลกอริทึม Simplex method

Ref. code: 25595809030157UGW

3.3 การหาความแปรปรวนของตัวประมาณขนาดประชากร

3.3.1 การหาความแปรปรวนของ $\hat{N}_{NewGT}$

จาก Pijitrattana และ Lerdsuwansri (2015) ได้นำเสนอความแปรปรวนของ $\hat{N}_{GT}$ ไว้ดังสมการ (2.10) ดังนี้

$$Var(\hat{N}_{GT}) = \frac{n\hat{p}_0}{(1-\hat{p}_0)^2} + \frac{n^2}{(1-\hat{p}_0)^4}\left(\frac{\hat{\kappa}}{\hat{\kappa}+1}\right)^2 f_1^{\left(\frac{\hat{\kappa}-1}{\hat{\kappa}+1}\right)} S^{-\left(\frac{2\hat{\kappa}}{\hat{\kappa}+1}\right)}\left\{\left(1-\frac{f_1}{\hat{N}}\right)+\hat{p}_0\right\}$$

โดยที่ $S=\sum_{i=1}^{m} if_i$, $\hat{p}_0 = \left(\frac{f_1}{S}\right)^{\frac{\hat{\kappa}}{\hat{\kappa}+1}}$ และ $\hat{\kappa}=\frac{\hat{\alpha}}{\hat{\beta}}$ ซึ่ง $\hat{\alpha}$ และ $\hat{\beta}$ เป็นค่าประมาณสัมประสิทธิ์การถดถอยจากสมการถดถอยเชิงเส้นแบบถ่วงน้ำหนักด้วยเมทริกซ์ทแยงมุมที่ประกอบด้วยความแปรปรวนผกผันของ $\hat{r}_i$; $w_i = \left[\hat{r}_i^2\left(\frac{1}{f_{i+1}}+\frac{1}{f_i}\right)\right]^{-1}$

ในทำนองเดียวกัน ความแปรปรวนของ $\hat{N}_{NewGT}$ คือ

$$Var(\hat{N}_{NewGT}) = \frac{n\hat{p}_0}{(1-\hat{p}_0)^2} + \frac{n^2}{(1-\hat{p}_0)^4}\left(\frac{\hat{\kappa}}{\hat{\kappa}+1}\right)^2 f_1^{\left(\frac{\hat{\kappa}-1}{\hat{\kappa}+1}\right)} S^{-\left(\frac{2\hat{\kappa}}{\hat{\kappa}+1}\right)}\left\{\left(1-\frac{f_1}{\hat{N}}\right)+\hat{p}_0\right\} \quad (3.5)$$

โดยที่ $S=\sum_{i=1}^{m} if_i$, $\hat{p}_0 = \left(\frac{f_1}{S}\right)^{\frac{\hat{\kappa}}{\hat{\kappa}+1}}$ และ $\hat{\kappa}=\frac{\hat{\alpha}}{\hat{\beta}}$ ซึ่ง $\hat{\alpha}$ และ $\hat{\beta}$ เป็นค่าประมาณสัมประสิทธิ์การถดถอยที่ได้จากการถดถอยควอนไทล์ที่ระดับ 0.50 โดยประมาณด้วยอัลกอริทึม Simplex method

3.3.2 การหาความแปรปรวนของ $\hat{N}_{NewLB}$

ความแปรปรวนของ $\hat{N}_{NewLB}$ หาได้โดยใช้ค่าคาดหมายแบบมีเงื่อนไข (Conditioning Expectation) โดยอาศัยแหล่งที่มาจาก 2 แหล่งคือจากความแปรปรวนเนื่องจากการประมาณค่า $\hat{f}_0$ และจากความแปรปรวนเนื่องจากขนาดของ n

ความแปรปรวนของตัวประมาณขนาดประชากร ที่แสดงในสมการ (2.20)

$$Var(\hat{N}) = Var_n\left\{E(\hat{N}\mid n)\right\} + E_n\left\{Var(\hat{N}\mid n)\right\}$$

โดยที่ E_n และ Var_n คือค่าเฉลี่ยและความแปรปรวนที่ขึ้นอยู่กับการแจกแจงของ n โดยที่ n มีการแจกแจงทวินาม (Binomial distribution) ที่มีพารามิเตอร์ N และ $1-p_0$

3.3.2.1 หา $Var_n\left\{E(\hat{N}|n)\right\}$

เนื่องจาก $E(\hat{N}|n) \approx n+\hat{f}_0$

ดังนั้น

$$Var_n\left\{E(\hat{N}|n)\right\} \approx Var_n\left\{n+\hat{f}_0\right\} \approx Var_n\left\{n\right\} \approx N(1-p_0)p_0 \quad (3.6)$$

จาก $p_0 = \dfrac{E(f_0)}{N} = \dfrac{f_0}{N}$

ดังนั้นจึงประมาณ $Var_n\left\{E(\hat{N}|n)\right\}$ ได้ดังนี้

$$Var_n\left\{E(\hat{N}|n)\right\} = \frac{nf_0}{(n+f_0)} = \frac{nf_1 e^{-(\hat{\alpha}+\hat{\beta})}}{(n+f_1 e^{-(\hat{\alpha}+\hat{\beta})})} \quad (3.7)$$

3.3.2.2 หา $E_n\left\{Var(\hat{N}|n)\right\}$

สมมติให้ $E_n\left\{Var(\hat{N}|n)\right\} \approx Var(\hat{N}|n)$

เนื่องจาก

$$Var(\hat{N}|n) = Var(\hat{f}_0) = Var(f_1 e^{-(\hat{\alpha}+\hat{\beta})})$$

ดังนั้น

$$E_n\left\{Var(\hat{N}|n)\right\} = Var(f_1 e^{-(\hat{\alpha}+\hat{\beta})})$$

ใช้วิธีค่าคาดหมายแบบมีเงื่อนไข (Conditioning Expectation) ในการหา $Var(f_1 e^{-(\hat{\alpha}+\hat{\beta})})$

จาก

$$Var(f_1 e^{-(\hat{\alpha}+\hat{\beta})}) = \underbrace{Var_{f_1}\left\{E(f_1 e^{-(\hat{\alpha}+\hat{\beta})}|f_1)\right\}}_{(1)} + \underbrace{E_{f_1}\left\{Var(f_1 e^{-(\hat{\alpha}+\hat{\beta})}|f_1)\right\}}_{(2)} \quad (3.8)$$

โดยที่ E_{f_1} และ Var_{f_1} คือค่าคาดหมายและความแปรปรวนของ f_1

โดยที่ f_1 มีการแจกแจงทวินามที่มีพารามิเตอร์ N และ p_1

Ref. code: 25595809030157UGW

พิจารณาส่วนที่ (1) ของสมการ (3.8)

โดยสมมติให้ $Var_{f_1}\left\{E(f_1 e^{-(\hat{\alpha}+\hat{\beta})} \mid f_1)\right\} \approx Var_{f_1}\left\{f_1 e^{-(\hat{\alpha}+\hat{\beta})}\right\}$

$$= e^{-(\hat{\alpha}+\hat{\beta})^2} Var_{f_1}(f_1)$$

$$= e^{-(\hat{\alpha}+\hat{\beta})2} Np_1(1-p_1)$$

จาก $p_1 = \frac{f_1}{N}$ จะได้

$$Var_{f_1}\left\{E(f_1 e^{-(\hat{\alpha}+\hat{\beta})} \mid f_1)\right\} = e^{-(\hat{\alpha}+\hat{\beta})2} f_1(1-\frac{f_1}{N}) \qquad (3.9)$$

และจากส่วนที่ (2) ของสมการ (3.8)

สมมติให้ $E_{f_1}\left\{Var(f_1 e^{-(\hat{\alpha}+\hat{\beta})} \mid f_1)\right\} \approx Var\left\{(f_1 e^{-(\hat{\alpha}+\hat{\beta})} \mid f_1)\right\}$

$$= f_1^2 Var(e^{-(\hat{\alpha}+\hat{\beta})}) \qquad (3.10)$$

และใช้วิธีเดลต้า (Delta method) เพื่อหา $Var(e^{-(\hat{\alpha}+\hat{\beta})})$

โดยกำหนดให้ $g(x) = e^{-x}$ และ $g'(x) = -e^{-x}$ จาก $Var(Y) \approx Var(x)[g'(\mu_x)]^2$ จะได้

$$Var(e^{-(\hat{\alpha}+\hat{\beta})}) \approx Var(\hat{\alpha}+\hat{\beta})(-e^{-(\hat{\alpha}+\hat{\beta})})^2 \approx (e^{-(\hat{\alpha}+\hat{\beta})})^2 Var(\hat{\alpha}+\hat{\beta}) \qquad (3.11)$$

แทนค่า $Var(e^{-(\hat{\alpha}+\hat{\beta})})$ จากสมการ (3.11) ในสมการ (3.10) จะได้

$$E_{f_1}\left\{Var(f_1 e^{-(\hat{\alpha}+\hat{\beta})} \mid f_1)\right\} \approx f_1^2 (e^{-(\hat{\alpha}+\hat{\beta})})^2 Var(\hat{\alpha}+\hat{\beta}) \qquad (3.12)$$

จากสมการ (3.7), (3.9) และ (3.12) แทนในสมการ (2.20) จะได้ค่าประมาณความแปรปรวนของ $\hat{N}_{NewLB}$ ดังนี้

$$Var(\hat{N}_{NewLB}) = \frac{nf_1 e^{-(\hat{\alpha}+\hat{\beta})}}{(n + f_1 e^{-(\hat{\alpha}+\hat{\beta})})} + e^{-(\hat{\alpha}+\hat{\beta})2} f_1(1-\frac{f_1}{\hat{N}}) + f_1^2 (e^{-(\hat{\alpha}+\hat{\beta})})^2 Var(\hat{\alpha}+\hat{\beta}) \qquad (3.13)$$

Ref. code: 25595809030157UGW

โดยที่ $Var(\hat{\alpha}+\hat{\beta})$ ประมาณจากการถดถอยควอนไทล์ที่ระดับ 0.50 โดยประมาณด้วยอัลกอริทึม Simplex method โดยข้อจำกัดของการศึกษาคือ $\hat{\alpha}$ และ $\hat{\beta}$ กำหนดให้เป็นอิสระต่อกัน

3.4 การหาช่วงเชื่อมั่นของขนาดประชากร

การสร้างช่วงเชื่อมั่นของขนาดประชากร จะกำหนดให้ตัวประมาณขนาดประชากร $\hat{N}$ มีการแจกแจงใกล้เคียงการแจกแจงปรกติ และในการสร้างช่วงเชื่อมั่น $(1-\alpha)100\%$ ของขนาดประชากร จะอาศัยวิธีของวาลด์ (Wald method) ซึ่งมีรูปแบบดังนี้

$$\hat{N} \pm z_{1-\frac{\alpha}{2}} \sqrt{Var(\hat{N})}$$

โดย $\hat{N}$ คือ ตัวประมาณขนาดประชากรแบบจุด

$z_{1-\frac{\alpha}{2}}$ คือ ควอนไทล์ที่ $1-\frac{\alpha}{2}$ ของการแจกแจงปรกติมาตรฐาน

$Var(\hat{N})$ คือ ความแปรปรวนของตัวประมาณขนาดประชากร

3.4.1 ช่วงเชื่อมั่นของขนาดประชากรโดยตัวประมาณ Generalized Turing $(\hat{N}_{GT})$

จากตัวประมาณขนาดประชากร Generalized Turing $(\hat{N}_{GT})$ ที่แสดงในสมการ (2.9)

$$\hat{N}_{GT} = \frac{n}{1-(\frac{f_1}{S})^{\frac{\hat{\kappa}}{\hat{\kappa}+1}}}$$

และค่าประมาณของ $Var(\hat{N}_{GT})$ (Pijitrattana & Lerdsuwansri,2015) ที่แสดงในสมการ (2.10)

$$Var(\hat{N}_{GT}) = \frac{n\hat{p}_0}{(1-\hat{p}_0)^2} + \frac{n^2}{(1-\hat{p}_0)^4}\left(\frac{\hat{\kappa}}{\hat{\kappa}+1}\right)^2 f_1^{\frac{\hat{\kappa}-1}{\hat{\kappa}+1}} S^{-\frac{2\hat{\kappa}}{\hat{\kappa}+1}} \left\{\left(1-\frac{f_1}{\hat{N}}\right)+\hat{p}_0\right\}$$

Ref. code: 25595809030157UGW

โดยที่ $S=\sum_{i=1}^{m} if_i$, $\hat{p}_0=\left(\frac{f_1}{S}\right)^{\frac{\hat{\kappa}}{\hat{\kappa}+1}}$ และ $\hat{\kappa}=\frac{\hat{\alpha}}{\hat{\beta}}$ โดยที่ $\hat{\alpha}$ และ $\hat{\beta}$ เป็นค่าประมาณสัมประสิทธิ์การถดถอยที่ได้จากสมการถดถอยเชิงเส้นแบบถ่วงน้ำหนักด้วยเมทริกซ์ทแยงมุมที่ประกอบด้วยความแปรปรวนผกผันของ $\hat{r}_i$ $;w_i=\left[\hat{r}_i^2\left(\frac{1}{f_{i+1}}+\frac{1}{f_i}\right)\right]^{-1}$ (Böhning et al.,2008)

ดังนั้นสร้างช่วงเชื่อมั่น $100(1-\alpha)\%$ ของขนาดประชากรโดยตัวประมาณขนาดประชากร Generalized Turing ได้ดังนี้

$$\left(\hat{N}_{GT}-z_{1-\frac{\alpha}{2}}\sqrt{Var(\hat{N}_{GT})},\hat{N}_{GT}+z_{1-\frac{\alpha}{2}}\sqrt{Var(\hat{N}_{GT})}\right) \quad (3.14)$$

3.4.2 ช่วงเชื่อมั่นของขนาดประชากรโดยตัวประมาณ Generalized Turing ที่ปรับปรุงโดยใช้การถดถอยควอนไทล์ที่ระดับ 0.50 $(\hat{N}_{NewGT})$

จากตัวประมาณขนาดประชากร Generalized Turing ที่ปรับปรุงโดยใช้การถดถอยควอนไทล์ที่ระดับ 0.50 $(\hat{N}_{NewGT})$ ซึ่งแสดงในสมการ (3.1)

$$\hat{N}_{NewGT}=\frac{n}{1-(\frac{f_1}{S})^{\frac{\hat{\kappa}}{\hat{\kappa}+1}}}$$

และค่าประมาณของ $Var(\hat{N}_{NewGT})$ แสดงในสมการ (3.5)

$$Var(\hat{N}_{NewGT})=\frac{n\hat{p}_0}{(1-\hat{p}_0)^2}+\frac{n^2}{(1-\hat{p}_0)^4}\left(\frac{\hat{\kappa}}{\hat{\kappa}+1}\right)^2 f_1^{\left(\frac{\hat{\kappa}-1}{\hat{\kappa}+1}\right)}S^{-\left(\frac{2\hat{\kappa}}{\hat{\kappa}+1}\right)}\left\{\left(1-\frac{f_1}{\hat{N}}\right)+\hat{p}_0\right\}$$

โดยที่ $\hat{\kappa}=\frac{\hat{\alpha}}{\hat{\beta}}$ และ $\hat{\alpha},\hat{\beta}$ เป็นค่าประมาณสัมประสิทธิ์การถดถอยที่ได้จากการถดถอยควอนไทล์ที่ระดับ 0.50 โดยประมาณด้วยอัลกอริทึม Simplex method

ดังนั้นสร้างช่วงเชื่อมั่น $100(1-\alpha)\%$ ของขนาดประชากรโดยตัวประมาณขนาดประชากร Generalized Turing ที่ปรับปรุงโดยใช้การถดถอยควอนไทล์ที่ระดับ 0.50 $(\hat{N}_{NewGT})$ ได้ดังนี้

Ref. code: 25595809030157UGW

$$\left(\hat{N}_{NewGT} - z_{1-\frac{\alpha}{2}}\sqrt{Var(\hat{N}_{NewGT})}, \hat{N}_{NewGT} + z_{1-\frac{\alpha}{2}}\sqrt{Var(\hat{N}_{NewGT})}\right) \quad (3.15)$$

3.4.3 ช่วงเชื่อมั่นของขนาดประชากรโดยตัวประมาณ Lanumteang&Böhning $(\hat{N}_{LB})$

จากตัวประมาณขนาดประชากร Lanumteang&Böhning ที่แสดงในสมการ (2.13)

$$\hat{N}_{LB} = n + \frac{3f_1^3 f_3}{4f_2^3}$$

และค่าประมาณของ $Var(\hat{N}_{LB})$ ที่แสดงในสมการ (2.14)

$$Var(\hat{N}_{LB}) = (\frac{9}{4})^2 \frac{f_1^5 f_3^2}{f_2^6}(\frac{f_1}{f_2}+1) + (\frac{3}{4})^2 \frac{f_1^6 f_3}{f_2^6}(1-\frac{f_3}{n}) + (\frac{3n}{4})\frac{f_1^3 f_3}{nf_2^3 + \frac{3}{4}f_1^3 f_3}$$

ดังนั้นสร้างช่วงเชื่อมั่น $100(1-\alpha)\%$ ของขนาดประชากรโดยตัวประมาณขนาดประชากร Lanumteang&Böhning ได้ดังนี้

$$\left(\hat{N}_{LB} - z_{1-\frac{\alpha}{2}}\sqrt{Var(\hat{N}_{LB})}, \hat{N}_{LB} + z_{1-\frac{\alpha}{2}}\sqrt{Var(\hat{N}_{LB})}\right) \quad (3.16)$$

3.4.4 ช่วงเชื่อมั่นของขนาดประชากรโดยตัวประมาณ Lanumteang&Böhning ที่ปรับปรุงโดยใช้การถดถอยควอนไทล์ที่ระดับ 0.50 $(\hat{N}_{NewLB})$

จากตัวประมาณขนาดประชากร Lanumteang&Böhning ที่ปรับปรุงโดยใช้การถดถอยควอนไทล์ที่ระดับ 0.50 $(\hat{N}_{NewLB})$ ที่แสดงในสมการ (3.4)

$$\hat{N}_{newLB} = n + \frac{f_1}{e^{\hat{\alpha}+\hat{\beta}}}$$

โดยที่ $\hat{\alpha}$ และ $\hat{\beta}$ เป็นค่าประมาณสัมประสิทธิ์การถดถอยที่ได้จากการถดถอยควอนไทล์ที่ระดับ 0.50 โดยประมาณด้วยอัลกอริทึม Simplex method

และค่าประมาณ $Var(\hat{N}_{NewLB})$ ความแปรปรวนของ $\hat{N}_{NewLB}$ ที่แสดงในสมการ (3.13)

Ref. code: 25595809030157UGW

$$Var(\hat{N}_{NewLB}) = \frac{nf_1 e^{-(\hat{\alpha}+\hat{\beta})}}{(n + f_1 e^{-(\hat{\alpha}+\hat{\beta})})} + e^{-(\hat{\alpha}+\hat{\beta})2} f_1(1-\frac{f_1}{\hat{N}}) + f_1^{\,2}(e^{-(\hat{\alpha}+\hat{\beta})})^2 Var(\hat{\alpha}+\hat{\beta})$$

โดยที่ $Var(\hat{\alpha}+\hat{\beta})$ ประมาณจากการถดถอยควอนไทล์ที่ระดับ 0.50 โดยประมาณด้วยอัลกอริทึม Simplex method โดยข้อจำกัดของการศึกษาคือ $\hat{\alpha}$ และ $\hat{\beta}$ กำหนดให้เป็นอิสระต่อกัน

ดังนั้นสร้างช่วงเชื่อมั่น $100(1-\alpha)\%$ ของขนาดประชากรโดยตัวประมาณขนาดประชากร Lanumteang&Böhning ที่ปรับปรุงโดยใช้การถดถอยควอนไทล์ที่ระดับ 0.50 ได้ดังนี้

$$\left(\hat{N}_{NewLB} - z_{1-\frac{\alpha}{2}}\sqrt{Var(\hat{N}_{NewLB})}, \hat{N}_{NewLB} + z_{1-\frac{\alpha}{2}}\sqrt{Var(\hat{N}_{NewLB})}\right) \quad (3.17)$$

3.5 การดำเนินการวิจัย

การดำเนินการวิจัยในงานวิจัยนี้ได้แบ่งออกเป็น 2 ส่วน ได้แก่

ส่วนที่ 1 เปรียบเทียบประสิทธิภาพของตัวประมาณขนาดประชากร Generalized Turing $(\hat{N}_{GT})$, ตัวประมาณขนาดประชากร Generalized Turing ที่ปรับปรุงโดยใช้การถดถอยควอนไทล์ที่ระดับ 0.50 $(\hat{N}_{NewGT})$, ตัวประมาณขนาดประชากร Lanumteang&Böhning $(\hat{N}_{LB})$ และตัวประมาณขนาดประชากร Lanumteang&Böhning ที่ปรับปรุงโดยใช้การถดถอยควอนไทล์ที่ระดับ 0.50 $(\hat{N}_{NewLB})$ ภายใต้สถานการณ์ต่างๆ

ส่วนที่ 2 เปรียบเทียบประสิทธิภาพของช่วงเชื่อมั่นของขนาดประชากรโดยตัวประมาณ $\hat{N}_{GT}, \hat{N}_{NewGT}, \hat{N}_{LB}$ และ $\hat{N}_{NewLB}$ ภายใต้สถานการณ์ต่างๆ

3.5.1 เปรียบเทียบประสิทธิภาพของตัวประมาณขนาดประชากร ภายใต้สถานการณ์ต่างๆ

3.5.1.1 กำหนดสถานการณ์จำลอง

3.5.1.1.1 กำหนดขนาดประชากร (N) ที่ศึกษาดังนี้ กรณีประชากรขนาดเล็ก $N = 500,\ 1{,}000$, ประชากรขนาดกลาง $N = 5{,}000,\ 10{,}000$ และประชากรขนาดใหญ่ $N = 50{,}000,\ 100{,}000$

3.5.1.1.2 กำหนดพารามิเตอร์

(1) การแจกแจงปัวซง (Poisson distribution) ที่มีพารามิเตอร์ λ คือ

$\lambda \in \{3,5,7\}$ รวมทั้งสิ้น 18 สถานการณ์

(2) การแจกแจงปัวซงแบบผสมที่มีพารามิเตอร์ λ_1, λ_2 กำหนดให้พารามิเตอร์ $(\lambda_1, \lambda_2) = (3,5)$ และ $(\lambda_1, \lambda_2) = (7,5)$ รวมทั้งสิ้น 12 สถานการณ์

(3) การแจกแจงทวินามเชิงลบ (Negative Binomial distribution) ที่มีพารามิเตอร์ π, κ คือ $\pi \in \{0.50, 0.80\}$ และ $\kappa \in \{4,9\}$ รวมทั้งสิ้น 24 สถานการณ์

3.5.1.2 ขั้นตอนการดำเนินการวิจัย

3.5.1.2.1 จำลองประชากรตามสถานการณ์ที่กำหนด

3.5.1.2.2 คำนวณค่า $f_{1,}, f_2, f_3, ..., f_m$, $S = \sum_{i=1}^{m} i f_i$ $, n = \sum_{i=1}^{m} f_i$ $, \hat{\kappa}$ $, \hat{\alpha}$ และ $\hat{\beta}$ ที่ใช้ในการคำนวณตัวประมาณขนาดประชากร

3.5.1.2.3 นำค่าที่คำนวณได้ในข้อ 3.5.1.2.2 มาคำนวณตัวประมาณขนาดประชากรทั้ง 4 ที่ศึกษา ดังนี้

(1) ตัวประมาณขนาดประชากร Generalized $(\hat{N}_{GT})$

$$\hat{N}_{GT} = \frac{n}{1 - (\frac{f_1}{S})^{\frac{\hat{\kappa}}{\hat{\kappa}+1}}}$$

โดยประมาณค่า $\hat{\kappa} = \frac{\hat{\alpha}}{\hat{\beta}}$ โดยที่ $\hat{\alpha}$ และ $\hat{\beta}$ เป็นค่าประมาณสัมประสิทธิ์การถดถอยที่ได้จากสมการถดถอยเชิงเส้นแบบถ่วงน้ำหนักด้วยเมทริกซ์ทแยงมุมที่ประกอบด้วยความแปรปรวนผกผันของ $\hat{r}_i$ $; w_i = \left[\hat{r}_i^2 \left(\frac{1}{f_{i+1}} + \frac{1}{f_i} \right) \right]^{-1}$ (Böhning et al.,2008)

(2) ตัวประมาณขนาดประชากร Generalized Turing ที่ปรับปรุงโดยใช้การถดถอยควอนไทล์ที่ระดับ 0.50 $(\hat{N}_{NewGT})$

$$\hat{N}_{NewGT} = \frac{n}{1 - (\frac{f_1}{S})^{\frac{\hat{\kappa}}{\hat{\kappa}+1}}}$$

โดยประมาณค่า $\hat{\kappa} = \frac{\hat{\alpha}}{\hat{\beta}}$ โดยที่ $\hat{\alpha}$ และ $\hat{\beta}$ เป็นค่าประมาณสัมประสิทธิ์การถดถอยที่ได้จากการถดถอยควอนไทล์ที่ระดับ 0.50

(3) ตัวประมาณขนาดประชากร Lanumteang&Böhning $(\hat{N}_{LB})$

$$\hat{N}_{LB} = n + \frac{3f_1^3 f_3}{4f_2^3}$$

(4) ตัวประมาณขนาดประชากร Lanumteang&Böhning ที่ปรับปรุงโดยใช้การถดถอยควอนไทล์ที่ระดับ 0.50 $(\hat{N}_{NewLB})$

$$\hat{N}_{NewLB} = n + \frac{f_1}{e^{\hat{\alpha}+\hat{\beta}}}$$

โดยที่ $\hat{\alpha}$ และ $\hat{\beta}$ เป็นค่าประมาณสัมประสิทธิ์การถดถอยที่ได้จากการถดถอยควอนไทล์ที่ระดับ 0.50

3.5.1.2.4 ทำซ้ำจากขั้นตอนที่ 3.5.1.2.1- 3.5.1.2.3 ทั้งหมด $B = 10,000$ รอบ

3.5.1.2.5 คำนวณค่าความเอนเอียงสัมพัทธ์ (Relative Bias) ค่าความแปรปรวนสัมพัทธ์ (Relative Variance) และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ (Relative Mean Square Error) ซึ่งคำนวณได้ดังนี้

(1) ค่าความเอนเอียงสัมพัทธ์ คือ

$$RBias = \frac{E(\hat{N}) - N}{N}$$

โดยที่ $\bar{\hat{N}} = \frac{1}{B}\sum_{i=1}^{B}\hat{N}_i$ เมื่อ B แทนจำนวนรอบของการวนซ้ำๆ ในการจำลอง

(2) ค่าความแปรปรวนสัมพัทธ์ คือ

$$RVar = \frac{1}{N^2}\sum_{i=1}^{B}\frac{(\hat{N}_i - \bar{\hat{N}})^2}{B}$$

(3) ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ คือ

$$RMSE = \frac{1}{N^2}\sum_{i=1}^{B}\frac{(\hat{N}_i - N)^2}{B}$$

สำหรับการจำลองสถานการณ์และการคำนวณทั้งหมดใช้โปรแกรม R Studio และทำซ้ำทั้งหมด $B = 10,000$ รอบ ซึ่งสามารถแสดงได้ดังภาพที่ 3.1

Ref. code: 25595809030157UGW

3.5.2 เปรียบเทียบประสิทธิภาพของช่วงเชื่อมั่นของขนาดประชากร ภายใต้สถานการณ์ต่างๆ

เปรียบเทียบประสิทธิภาพของช่วงเชื่อมั่นของขนาดประชากรโดยตัวประมาณ $\hat{N}_{GT}, \hat{N}_{NewGT}, \hat{N}_{LB}$ และ $\hat{N}_{NewLB}$ ภายใต้สถานการณ์ต่างๆ แสดงขั้นตอนได้ดังนี้

1. จำลองข้อมูลภายใต้สถานการณ์ต่างๆ โดยใช้โปรแกรม R Studio

2. คำนวณค่า $f_0, f_1, f_2, ..., f_m$ โดยที่ m คือ จำนวนครั้งสูงสุดที่ถูกพบ

3. คำนวณค่า $\hat{r}_i = \dfrac{(i+1)f_{i+1}}{f_i}$ และ $w_i = \left[\hat{r}_i^2\left(\dfrac{1}{f_{i+1}} + \dfrac{1}{f_i}\right)\right]^{-1}$

4. คำนวณค่า $\hat{\alpha}, \hat{\beta}$ โดยใช้การถดถอยเชิงเส้นอย่างง่ายแบบถ่วงน้ำหนัก และใช้การถดถอยควอนไทล์ที่ระดับ 0.50

5. คำนวณค่า $\hat{\kappa}$

6. คำนวณค่า $\hat{N}_{GT}$ จากสมการ (2.9) , $\hat{N}_{NewGT}$ จากสมการ (3.1) , $\hat{N}_{LB}$ จากสมการ (2.13) และ $\hat{N}_{NewLB}$ จากสมการ (3.4)

7. คำนวณค่า $Var(\hat{N}_{GT})$ จากสมการ (2.10) , $Var(\hat{N}_{NewGT})$ จากสมการ (3.5) , $Var(\hat{N}_{LB})$ จากสมการ (2.14) และ $Var(\hat{N}_{NewLB})$ จากสมการ (3.13)

8. คำนวณช่วงเชื่อมั่นของขนาดประชากรโดยตัวประมาณ $\hat{N}_{GT}, \hat{N}_{NewGT}, \hat{N}_{LB}$ และ $\hat{N}_{NewLB}$ จากที่ได้แสดงไว้ในข้างต้น

9. พิจารณาว่าช่วงเชื่อมั่นที่ได้ว่าครอบคลุมค่าของขนาดประชากร N หรือไม่ ถ้าช่วงเชื่อมั่นที่ได้ครอบคลุมค่า N ให้ค่าเป็น 1 และถ้าช่วงเชื่อมั่นที่ได้ไม่ครอบคลุมค่า N ให้ค่าเป็น 0

10. ทำขั้นตอนที่ 1-9 ซ้ำ จำนวน $B = 10,000$ รอบ นับจำนวนของช่วงเชื่อมั่นที่ครอบคลุมขนาดประชากร N แล้วบวกสะสมไว้ นำไปคำนวณหาค่าประมาณสัมประสิทธิ์ความเชื่อมั่นโดยนำจำนวนช่วงเชื่อมั่นที่ครอบคลุมทั้งหมด หารด้วยจำนวนรอบที่ทำซ้ำ

11. เปรียบเทียบค่าประมาณสัมประสิทธิ์ความเชื่อมั่นกับสัมประสิทธิ์ความเชื่อมั่นที่กำหนด

เกณฑ์ที่ใช้เปรียบเทียบวิธีการประมาณค่าแบบช่วงที่ระดับความเชื่อมั่นที่ 95% คือ ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นต้องมีค่าไม่ต่ำกว่า 0.9464 จะถือว่าช่วงเชื่อมั่นที่สร้างจากตัวประมาณขนาดประชากรนั้นมีค่าประมาณสัมประสิทธิ์ความเชื่อมั่นไม่ต่ำกว่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด จากนั้นคำนวณค่าความกว้างเฉลี่ยของช่วงเชื่อมั่น (Average Length) เฉพาะกรณีที่ช่วงเชื่อมั่นที่สร้างจากตัวประมาณขนาดประชากรที่ให้ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นไม่ต่ำกว่า

Ref. code: 25595809030157UGW

สัมประสิทธิ์ความเชื่อมั่นที่กำหนด โดยคำนวณจากผลต่างของขีดจำกัดบนและขีดจำกัดล่าง หารด้วยจำนวนรอบที่ทำซ้ำ

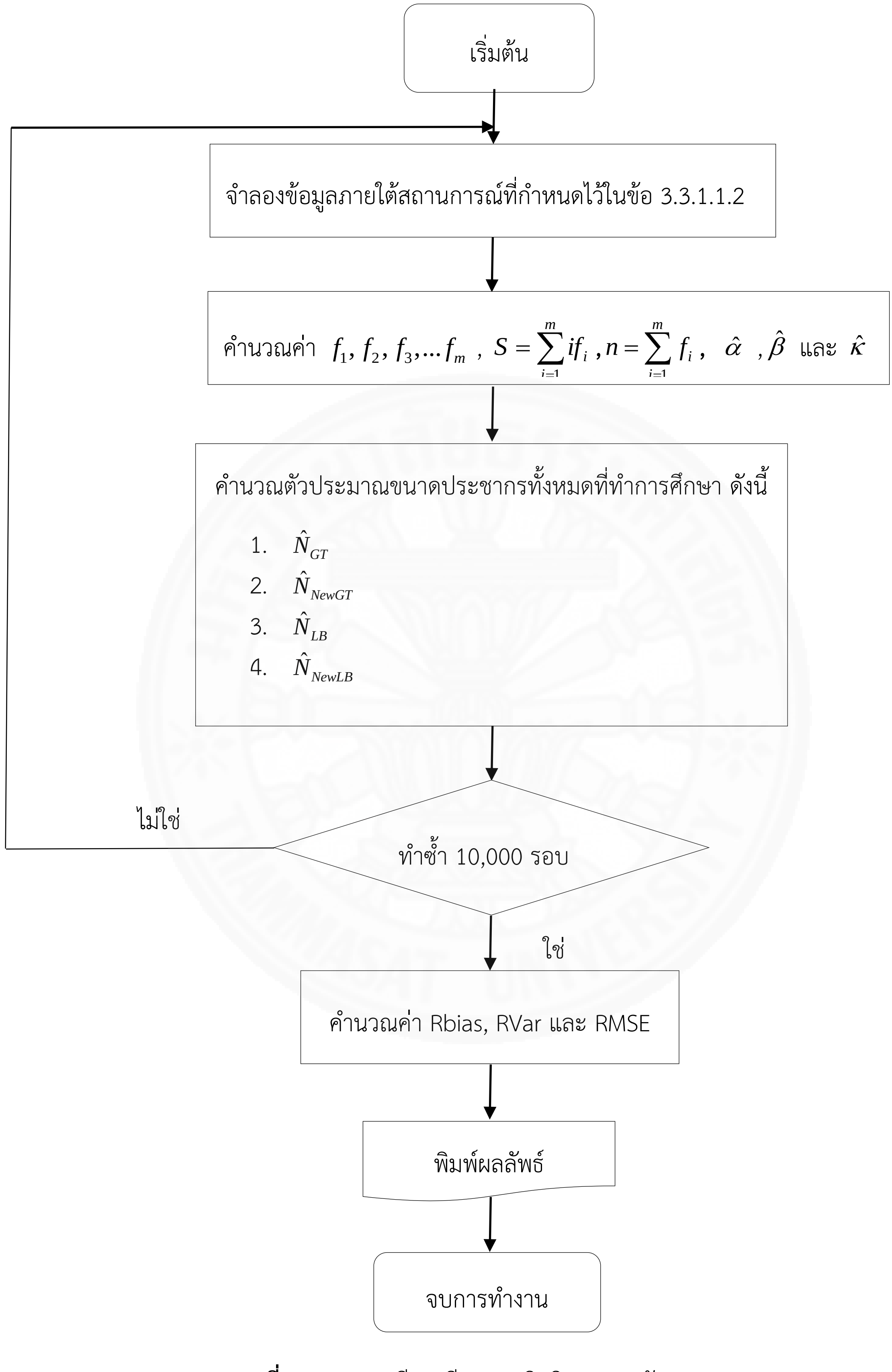
12. เปรียบเทียบประสิทธิภาพของช่วงเชื่อมั่น โดยพิจารณาจาก 2 ส่วนคือ

(1) ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นที่มีค่าไม่ต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นกำหนด

(2) ความกว้างเฉลี่ยของช่วงเชื่อมั่นมีค่าน้อยที่สุด

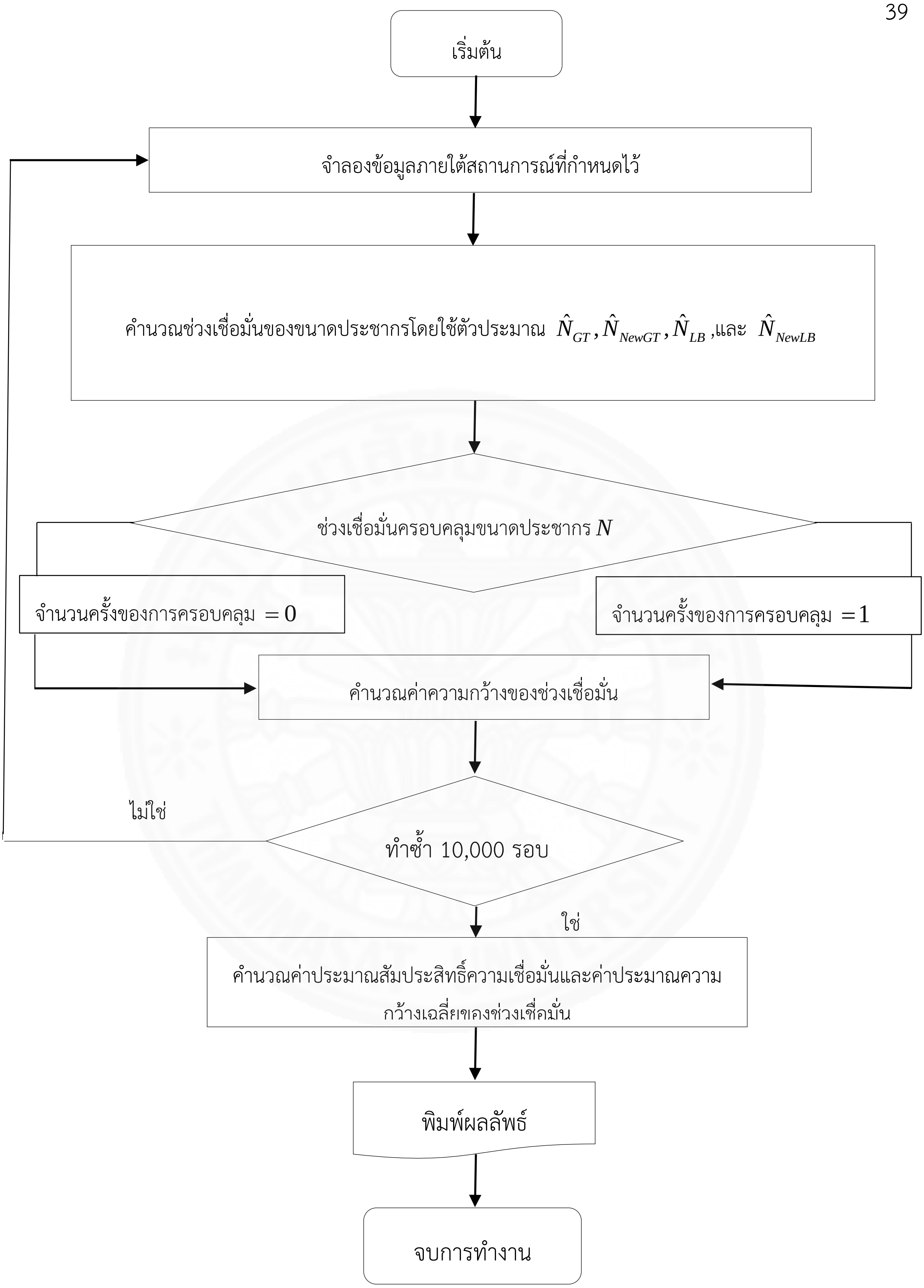
13. สรุปผลการวิจัย ซึ่งแสดงได้ดังภาพที่ 3.2

Ref. code: 25595809030157UGW



ภาพที่ 3.1 แสดงเปรียบเทียบประสิทธิภาพของตัวประมาณขนาดประชากร

Ref. code: 25595809030157UGW



ภาพที่ 3.2 แสดงเปรียบเทียบประสิทธิภาพช่วงเชื่อมั่นของขนาดประชากร

บทที่ 4
ผลการวิจัย

การวิจัยนี้เป็นการศึกษาโดยแบ่งออกเป็น 2 ส่วน ได้แก่ ส่วนที่ 1 เปรียบเทียบประสิทธิภาพของตัวประมาณขนาดประชากรทั้ง 4 คือ $\hat{N}_{GT}, \hat{N}_{NewGT}, \hat{N}_{LB}$ และ $\hat{N}_{NewLB}$ เกณฑ์ที่ใช้ในการเปรียบเทียบคือ ค่าความเอนเอียงสัมพัทธ์ (Relative Bias), ค่าความแปรปรวนสัมพัทธ์ (Relative Variance) และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ (Relative Mean Square Error) โดยตัวประมาณใดที่ให้ค่าความเอนเอียงสัมพัทธ์, ค่าความแปรปรวนสัมพัทธ์และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ต่ำสุด ถือว่ามีประสิทธิภาพสูงสุด และส่วนที่ 2 เปรียบเทียบประสิทธิภาพของช่วงเชื่อมั่นของขนาดประชากรที่สร้างจากตัวประมาณขนาดประชากรทั้ง 4 คือ $\hat{N}_{GT}, \hat{N}_{NewGT}, \hat{N}_{LB}$ และ $\hat{N}_{NewLB}$ เกณฑ์ที่ใช้ในการเปรียบเทียบคือ ค่าประมาณสัมประสิทธิ์ความเชื่อมั่น $(1-\hat{\alpha})$ และค่าประมาณความกว้างเฉลี่ยของช่วงเชื่อมั่น (Averagre Length) โดยช่วงเชื่อมั่นที่สร้างจากตัวประมาณขนาดประชากรใดให้ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นไม่ต่ำกว่า 0.9464 จากนั้นจะพิจารณาค่าประมาณความกว้างเฉลี่ยของช่วงเชื่อมั่นต่อไป ถ้าช่วงเชื่อมั่นที่สร้างจากตัวประมาณขนาดประชากรใดให้ค่าประมาณความกว้างเฉลี่ยของช่วงเชื่อมั่นแคบที่สุด ถือว่าช่วงเชื่อมั่นที่สร้างจากตัวประมาณขนาดประชากรนั้นมีประสิทธิภาพสูงสุด โดยผลการวิจัยในบทที่ 4 นี้จะนำเสนอผลที่ได้จากการจำลองสถานการณ์พร้อมทั้งการอธิบายผลที่ได้ นอกจากนี้ยังนำตัวประมาณขนาดประชากรที่ศึกษาทั้ง 4 มาประยุกต์ใช้กับข้อมูลจริงด้วย โดยผลการวิจัยแสดงได้ดังนี้

การนำเสนอผลการวิจัยนำเสนอในรูปแบบตารางเพื่อให้เข้าใจชัดเจน และใช้สัญลักษณ์แทนความหมายต่างๆ ในตารางดังต่อไปนี้

GT หมายถึง ตัวประมาณขนาดประชากร Generalized Turing

NewGT หมายถึง ตัวประมาณขนาดประชากร Generalized Turing ที่ปรับปรุงโดยใช้การถดถอยควอนไทล์ที่ระดับ 0.50

LB หมายถึง ตัวประมาณขนาดประชากร Lanumteang&Böhning

NewLB หมายถึง ตัวประมาณขนาดประชากร Lanumteang&Böhning ปรับปรุงโดยใช้การถดถอยควอนไทล์ที่ระดับ 0.50

$\bar{\hat{N}}$ หมายถึง ค่าเฉลี่ยของตัวประมาณขนาดประชากรจากการจำลองสถานการณ์

Ref. code: 25595809030157UGW

RBias หมายถึง ค่าความเอนเอียงสัมพัทธ์

RVar หมายถึง ค่าความแปรปรวนสัมพัทธ์

RMSE หมายถึง ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์

$1-\hat{\alpha}$ หมายถึง ค่าประมาณสัมประสิทธิ์ความเชื่อมั่น

Average Length หมายถึง ค่าประมาณความกว้างเฉลี่ยของช่วงเชื่อมั่น

4.1 เปรียบเทียบประสิทธิภาพของตัวประมาณขนาดประชากร

ผลจากการจำลองสถานการณ์ได้แสดงไว้ดังตารางที่ 4.1-4.9 ซึ่งแบ่งออกเป็น 3 ส่วน คือ ผลจากการจำลองภายใต้ตัวแบบการแจกแจงปัวซงที่มีพารามิเตอร์ λ ผลจากการจำลองภายใต้ตัวแบบการแจกแจงปัวซงแบบผสมที่มีพารามิเตอร์ λ_1, λ_2 และผลจากการจำลองภายใต้ตัวแบบการแจกแจงทวินามเชิงลบที่มีพารามิเตอร์ κ, π โดยพิจารณาจากค่าความเอนเอียงสัมพัทธ์ ค่าความแปรปรวนสัมพัทธ์ และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ สามารถอธิบายผลการจำลองได้ดังนี้

4.1.1 การจำลองภายใต้ตัวแบบการแจกแจงปัวซงที่มีพารามิเตอร์ λ

จากการจำลองภายใต้ตัวแบบการแจกแจงปัวซงที่มีพารามิเตอร์ λ โดยกำหนดสถานการณ์ให้พารามิเตอร์ $\lambda = 3, 5, 7$ และขนาดประชากร $N = 500, 1{,}000, 5{,}000, 10{,}000, 50{,}000$ และ $100{,}000$ ตามลำดับ รวมทั้งสิ้น 18 สถานการณ์ แสดงผลลัพธ์ได้ดังตารางที่ 4.1-4.3

4.1.1.1 พิจารณาค่าความเอนเอียงสัมพัทธ์

จากตารางที่ 4.1-4.3 จะเห็นได้ว่า ตัวประมาณขนาดประชากรทั้ง 4 ให้ค่าความเอนเอียงสัมพัทธ์ ($RBias$) เป็นบวกในทุกๆ สถานการณ์ กล่าวได้ว่า ตัวประมาณขนาดประชากรทั้ง 4 มีค่า overestimate ในทุกๆ สถานการณ์

เมื่อพิจารณา $\lambda = 3$ พบว่า $\hat{N}_{NewLB}$ ให้ค่า $RBias$ ต่ำที่สุดในทุกๆ ขนาดประชากร และเมื่อประชากรมีขนาดใหญ่ พบว่า $\hat{N}_{NewGT}, \hat{N}_{LB}$ และ $\hat{N}_{NewLB}$ จะให้ค่า $RBias$ เข้าใกล้ 0 มากขึ้น กล่าวได้ว่า ตัวประมาณนั้นเป็น asymptotic unbiased estimator และเมื่อเปรียบเทียบประสิทธิภาพของค่า $RBias$ ระหว่าง $\hat{N}_{GT}$ และ $\hat{N}_{NewGT}$ พบว่า $\hat{N}_{GT}$ ให้ค่า $RBias$ ต่ำกว่าของ $\hat{N}_{NewGT}$ ในทุกๆ ขนาดประชากร และเมื่อเปรียบเทียบประสิทธิภาพของค่า $RBias$ ระหว่าง $\hat{N}_{LB}$ และ $\hat{N}_{NewLB}$ พบว่า $\hat{N}_{NewLB}$ ให้ค่า $\hat{N}_{NewLB}$ ต่ำกว่าของ $\hat{N}_{LB}$ ในทุกๆ ขนาดประชากร

Ref. code: 25595809030157UGW

เมื่อพิจารณา $\lambda = 5$ พบว่า $\hat{N}_{NewLB}$ มีค่า $RBias$ ต่ำที่สุด ในทุกๆ ขนาดประชากร ส่วน $\hat{N}_{LB}$ มีค่า $RBias$ สูงที่สุดเมื่อประชากรมีขนาดเล็ก ($N = 500,1,000$) แต่เมื่อประชากรมีขนาดใหญ่ขึ้น พบว่า $\hat{N}_{GT}$ มีค่า $RBias$ สูงที่สุด และเมื่อประชากรมีขนาดใหญ่ ตัวประมาณขนาดประชากรทั้ง 4 ให้ค่า $RBias$ เข้าใกล้ 0 หรือเรียกว่าเป็น asymptotic unbiased estimator โดยเฉพาะ $\hat{N}_{NewLB}$ และเมื่อเปรียบเทียบประสิทธิภาพของค่า $RBias$ ระหว่าง $\hat{N}_{GT}$ และ $\hat{N}_{NewGT}$ พบว่า $\hat{N}_{NewGT}$ ให้ค่า $RBias$ ต่ำกว่าของ $\hat{N}_{GT}$ เมื่อประชากรมีขนาดใหญ่ และเมื่อเปรียบเทียบประสิทธิภาพของค่า $RBias$ ระหว่าง $\hat{N}_{LB}$ และ $\hat{N}_{NewLB}$ พบว่า $\hat{N}_{NewLB}$ ให้ค่า $\hat{N}_{NewLB}$ ต่ำกว่าของ $\hat{N}_{LB}$ ในทุกๆ ขนาดประชากร

เมื่อพิจารณา $\lambda = 7$ พบว่า $\hat{N}_{NewLB}$ มีค่า $RBias$ ต่ำที่สุดในทุกๆ ขนาดประชากร และ $\hat{N}_{LB}$ ค่อนข้างสูงเกือบทุกขนาดประชากร และเมื่อประชากรมีขนาดใหญ่ ตัวประมาณขนาดประชากรทั้ง 4 ให้ค่า $RBias$ เข้าใกล้ 0 หรือเป็น asymptotic unbiased estimator โดยเฉพาะ $\hat{N}_{NewLB}$ และเมื่อเปรียบเทียบประสิทธิภาพของค่า $RBias$ ระหว่าง $\hat{N}_{GT}$ และ $\hat{N}_{NewGT}$ พบว่า $\hat{N}_{NewGT}$ ให้ค่า $RBias$ ต่ำกว่าของ $\hat{N}_{GT}$ เมื่อประชากรมีขนาดใหญ่ และเมื่อเปรียบเทียบประสิทธิภาพของค่า $RBias$ ระหว่าง $\hat{N}_{LB}$ และ $\hat{N}_{NewLB}$ พบว่า $\hat{N}_{NewLB}$ ให้ค่า $\hat{N}_{NewLB}$ ต่ำกว่าของ $\hat{N}_{LB}$ ในทุกๆ ขนาดประชากร

4.1.1.2 พิจารณาค่าความแปรปรวนสัมพัทธ์

จากตารางที่ 4.1-4.3 เมื่อพิจารณา $\lambda = 3$ พบว่า $\hat{N}_{GT}$ มีค่าความแปรปรวนสัมพัทธ์ ($RVar$) ต่ำที่สุด และพบว่า $\hat{N}_{LB}$ มีค่า $RVar$ สูงที่สุด ในทุกๆ สถานการณ์ และเมื่อเปรียบเทียบประสิทธิภาพของ $\hat{N}_{GT}$ กับ $\hat{N}_{NewGT}$ พบว่า $\hat{N}_{GT}$ ให้ค่า $RVar$ ต่ำกว่า $\hat{N}_{NewGT}$ ในทุกๆ สถานการณ์ และเมื่อเปรียบเทียบประสิทธิภาพของ $\hat{N}_{LB}$ และ $\hat{N}_{NewLB}$ พบว่า ต่ำกว่า $\hat{N}_{LB}$ ในทุกๆ ขนาดประชากร

เมื่อพิจารณา $\lambda = 5$ พบว่า $\hat{N}_{GT}$ มีค่า $RVar$ ต่ำที่สุดเกือบทุกสถานการณ์ ยกเว้นเมื่อประชากรมีขนาดเล็ก ($N = 500$) พบว่า $\hat{N}_{NewLB}$ ให้ค่า $RVar$ ต่ำที่สุด และพบว่า $\hat{N}_{LB}$ มีค่า $RVar$ สูงที่สุด ในทุกๆ สถานการณ์ และเมื่อเปรียบเทียบประสิทธิภาพของ $\hat{N}_{GT}$ กับ $\hat{N}_{NewGT}$ พบว่า $\hat{N}_{GT}$ ให้ค่า $RVar$ ต่ำกว่า $\hat{N}_{NewGT}$ ในทุกๆสถานการณ์ และเมื่อเปรียบเทียบประสิทธิภาพของ $\hat{N}_{LB}$ และ $\hat{N}_{NewLB}$ พบว่า $\hat{N}_{NewLB}$ มีค่า $RVar$ ต่ำกว่า $\hat{N}_{LB}$ ในทุกๆ ขนาดประชากร

เมื่อพิจารณา $\lambda = 7$ พบว่า $\hat{N}_{GT}$ มีค่า $RVar$ ต่ำที่สุดเกือบทุก สถานการณ์ แต่เมื่อประชากรมีขนาดใหญ่ พบว่า $\hat{N}_{NewLB}$ ให้ค่า $RVar$ ต่ำใกล้เคียงกับค่า $RVar$ ของ $\hat{N}_{GT}$ และค่า $RVar$ ของตัวประมาณทั้ง 4 ให้ค่าใกล้เคียงกันมากและเข้าใกล้ 0 ในทุกๆ ขนาดประชากร และเมื่อ

Ref. code: 25595809030157UGW

เปรียบเทียบประสิทธิภาพของ $\hat{N}_{GT}$ กับ $\hat{N}_{NewGT}$ พบว่า $\hat{N}_{GT}$ ให้ค่า $RVar$ ใกล้เคียงกันในทุกๆ ขนาดประชากร ผู้วิจัยจึงสรุปว่า ประสิทธิภาพของตัวประมาณ $\hat{N}_{GT}$ และ $\hat{N}_{NewGT}$ พิจารณาจากค่า $RVar$ นั้นให้ประสิทธิภาพพอๆ กัน และเมื่อเปรียบเทียบประสิทธิภาพของ $\hat{N}_{LB}$ และ $\hat{N}_{NewLB}$ พบว่า $\hat{N}_{NewLB}$ มีค่า $RVar$ ต่ำกว่า $\hat{N}_{LB}$ ในทุกๆ ขนาดประชากร

4.1.1.3 พิจารณาค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์

จากตารางที่ 4.1-4.3 เมื่อพิจารณาพารามิเตอร์ $\lambda = 3,5,7$ พบว่า $\hat{N}_{LB}$ มีค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ ($RMSE$) สูงกว่าตัวประมาณอื่นๆ อย่างเห็นได้ชัดเจน ในทุกๆ สถานการณ์ และเมื่อพิจารณา $\lambda = 3$ พบว่า $\hat{N}_{GT}$ มีค่า $RMSE$ ใกล้เคียงกับ $\hat{N}_{NewLB}$ และมีค่าต่ำที่สุดในทุกๆ ขนาดประชากร และเมื่อเปรียบเทียบประสิทธิภาพของ $\hat{N}_{GT}$ กับ $\hat{N}_{NewGT}$ พบว่า $\hat{N}_{GT}$ ให้ค่า $RMSE$ ต่ำกว่า $\hat{N}_{NewGT}$ ในทุกๆ ขนาดประชากร และเมื่อเปรียบเทียบประสิทธิภาพของ $\hat{N}_{LB}$ และ $\hat{N}_{NewLB}$ พบว่า $\hat{N}_{NewLB}$ มีค่า $RVar$ ต่ำกว่า $\hat{N}_{LB}$ ในทุกๆ ขนาดประชากร

เมื่อพิจารณา $\lambda = 5$ พบว่า $\hat{N}_{NewLB}$ ให้ค่า $RMSE$ ต่ำที่สุดเกือบทุกสถานการณ์ ยกเว้นเมื่อประชากรมีขนาด $N = 50,000$ พบว่า $\hat{N}_{GT}$ ให้ค่า $RMSE$ ต่ำที่สุด และเมื่อเปรียบเทียบประสิทธิภาพของ $\hat{N}_{GT}$ กับ $\hat{N}_{NewGT}$ พบว่า โดยภาพรวม $\hat{N}_{GT}$ ให้ค่า $RMSE$ ต่ำกว่าของ $\hat{N}_{NewGT}$ และเมื่อเปรียบเทียบประสิทธิภาพของ $\hat{N}_{LB}$ และ $\hat{N}_{NewLB}$ พบว่า $\hat{N}_{NewLB}$ มีค่า $RVar$ ต่ำกว่า $\hat{N}_{LB}$ ในทุกๆ ขนาดประชากร

เมื่อพิจารณา $\lambda = 7$ พบว่า $\hat{N}_{NewLB}$ มีค่า $RMSE$ ต่ำที่สุดเกือบทุกสถานการณ์ และเมื่อเปรียบเทียบประสิทธิภาพของ $\hat{N}_{GT}$ กับ $\hat{N}_{NewGT}$ พบว่า $\hat{N}_{GT}$ และ $\hat{N}_{NewGT}$ ให้ค่า $RMSE$ ใกล้เคียงกันเมื่อประชากรมีขนาดใหญ่ขึ้น และเมื่อเปรียบเทียบประสิทธิภาพของ $\hat{N}_{LB}$ และ $\hat{N}_{NewLB}$ พบว่า $\hat{N}_{NewLB}$ มีค่า $RVar$ ต่ำกว่า $\hat{N}_{LB}$ ในทุกๆ ขนาดประชากร

Ref. code: 25595809030157UGW

4.1.2 การจำลองภายใต้ตัวแบบการแจกแจงปัวซงแบบผสมที่มีพารามิเตอร์ λ_1, λ_2

จากการจำลองภายใต้ตัวแบบการแจกแจงปัวซงแบบผสมที่มีพารามิเตอร์ λ_1, λ_2 โดยกำหนดสถานการณ์ให้พารามิเตอร์ $(\lambda_1, \lambda_2) = (3,5)$ และ $(\lambda_1, \lambda_2) = (7,5)$ สำหรับ $N = 500, 1,000, 5,000, 10,000, 50,000, 100,000$ ตามลำดับ รวมทั้งสิ้น 12 สถานการณ์ แสดงผลลัพธ์ได้ดังตารางที่ 4.4-4.5

4.1.2.1 พิจารณาค่าความเอนเอียงสัมพัทธ์

จากตารางที่ 4.4-4.5 พิจารณากรณีพารามิเตอร์ $(\lambda_1, \lambda_2) = (3,5)$ จะเห็นว่า เมื่อประชากรมีขนาดเล็ก ($N = 500, 1,000$) พบว่าค่า $RBias$ ของ $\hat{N}_{GT}$ มีค่าติดลบ เนื่องจากค่าประมาณที่ได้ของ $\hat{N}_{GT}$ นั้นต่ำกว่าค่าจริงที่กำหนดหรือเรียกว่า underestimate และเมื่อประชากรมีขนาดใหญ่ขึ้น พบว่าค่า $\hat{N}_{NewLB}$ ให้ค่าเป็นบวก และเมื่อเปรียบเทียบค่า $RBias$ ของ $\hat{N}_{GT}$ และ $\hat{N}_{NewGT}$ พบว่า ค่า $RBias$ ของ $\hat{N}_{GT}$ ต่ำกว่าของ $\hat{N}_{NewGT}$ ทุกๆ สถานการณ์ ส่วนค่า $RBias$ ของ $\hat{N}_{LB}$ และ $\hat{N}_{NewLB}$ พบว่า ค่า $RBias$ ของ $\hat{N}_{NewLB}$ ต่ำกว่าของ $\hat{N}_{LB}$ เมื่อขนาดประชากรขนาดเล็ก แต่เมื่อประชากรมีขนาดใหญ่ พบว่าค่า $RBias$ ของ $\hat{N}_{LB}$ จะต่ำกว่าของ $\hat{N}_{NewLB}$

พิจารณากรณีพารามิเตอร์ $(\lambda_1, \lambda_2) = (7,5)$ พบว่า ตัวประมาณขนาดประชากรทั้ง 4 ให้ค่าเป็นบวกในทุกๆ สถานการณ์หรือเรียกว่า overestimate และค่า $RBias$ ของ $\hat{N}_{GT}$ สูงขึ้น เมื่อขนาดประชากรใหญ่ขึ้น ในทางตรงกันข้าม พบว่าค่า $RBias$ ของ $\hat{N}_{LB}, \hat{N}_{NewLB}$ จะลดลง เมื่อประชากรใหญ่ขึ้นหรือมีคุณสมบัติ asymptotic unbiased estimator และค่า $RBias$ ของ $\hat{N}_{NewLB}$ ต่ำที่สุดเกือบทุกสถานการณ์ ยกเว้นเมื่อ $N = 500$ ค่า $RBias$ ของ $\hat{N}_{GT}$ ให้ค่าต่ำที่สุด และเมื่อเปรียบเทียบประสิทธิภาพจากค่า $RBias$ ของ $\hat{N}_{GT}$ และ $\hat{N}_{NewGT}$ พบว่า ค่า $RBias$ ของ $\hat{N}_{GT}$ ต่ำกว่าของ $\hat{N}_{NewGT}$ ทุกๆ สถานการณ์ และระหว่าง $\hat{N}_{LB}$ และ $\hat{N}_{NewLB}$ พบว่า ค่า $RBias$ ของ $\hat{N}_{NewLB}$ ต่ำกว่าของ $\hat{N}_{LB}$ ในทุกๆ สถานการณ์

4.1.2.2 พิจารณาค่าความแปรปรวนสัมพัทธ์

จากตาราง 4.4-4.5 พิจารณากรณีพารามิเตอร์ $(\lambda_1, \lambda_2) = (3,5)$ พบว่า ค่า $RVar$ ของ $\hat{N}_{GT}$ และ $\hat{N}_{NewLB}$ มีค่าใกล้เคียงกันมากในทุกๆ ขนาดประชากร และพบว่าเมื่อเปรียบเทียบประสิทธิภาพของค่า $RVar$ ของ $\hat{N}_{GT}$ และ $\hat{N}_{NewGT}$ พบว่าค่า $RVar$ ของ $\hat{N}_{GT}$ มีค่าต่ำกว่าของ $\hat{N}_{NewGT}$ ในทุกๆ สถานการณ์ และเมื่อเปรียบเทียบประสิทธิภาพของ $\hat{N}_{LB}$ และ $\hat{N}_{NewLB}$ พบว่าค่า $RVar$ ของ $\hat{N}_{NewLB}$ มีค่าต่ำกว่าของ $\hat{N}_{LB}$ มากในทุกๆ สถานการณ์

Ref. code: 25595809030157UGW

พิจารณากรณีพารามิเตอร์ $(\lambda_1,\lambda_2)=(7,5)$ พบว่า ค่า $RVar$ ของตัวประมาณขนาดประชากรทั้ง 4 ให้ค่าเข้าใกล้ 0 เมื่อประชากรมีขนาดใหญ่ขึ้นหรือมี และพบว่าค่า $RVar$ ของ $\hat{N}_{NewLB}$ ต่ำที่สุด เกือบทุกสถานการณ์ ยกเว้นเมื่อ $N=100{,}000$ พบว่าค่า $RVar$ ของ $\hat{N}_{GT}$ ให้ค่าต่ำที่สุด ประชากร และเมื่อเปรียบเทียบประสิทธิภาพของค่า $RVar$ ของ $\hat{N}_{GT}$ และ $\hat{N}_{NewGT}$ พบว่าค่า $RVar$ ของ $\hat{N}_{GT}$ มีค่าต่ำกว่าของ $\hat{N}_{NewGT}$ ในทุกๆ สถานการณ์ และเมื่อเปรียบเทียบประสิทธิภาพของ $\hat{N}_{LB}$ และ $\hat{N}_{NewLB}$ พบว่าค่า $RVar$ ของ $\hat{N}_{NewLB}$ มีค่าต่ำกว่าของ $\hat{N}_{LB}$ ในทุกๆ สถานการณ์

4.1.2.3 พิจารณาค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์

จากตาราง 4.4-4.5 พิจารณากรณีพารามิเตอร์ $(\lambda_1,\lambda_2)=(3,5)$ พบว่า ค่า $RMSE$ ของ $\hat{N}_{GT}$ มีค่าต่ำที่สุดในทุกๆ ขนาดประชากร และเมื่อเปรียบเทียบประสิทธิภาพของค่า $RMSE$ ของ $\hat{N}_{GT}$ และ $\hat{N}_{NewGT}$ พบว่าค่า $RMSE$ ของ $\hat{N}_{GT}$ มีค่าต่ำกว่าของ $\hat{N}_{NewGT}$ ในทุกๆ สถานการณ์ และเมื่อเปรียบเทียบประสิทธิภาพของ $\hat{N}_{LB}$ และ $\hat{N}_{NewLB}$ พบว่าค่า $RMSE$ ของ $\hat{N}_{NewLB}$ มีค่าต่ำกว่าของ $\hat{N}_{LB}$ ในทุกๆ สถานการณ์

พิจารณากรณีพารามิเตอร์ $(\lambda_1,\lambda_2)=(7,5)$ พบว่า ค่า $RMSE$ ของตัวประมาณขนาดประชากรทั้ง 4 ให้ค่าเข้าใกล้ 0 เมื่อประชากรมีขนาดใหญ่ขึ้น หรือเป็น asymptotic unbiased estimator และค่า $RMSE$ ของ $\hat{N}_{NewLB}$ มีค่าต่ำที่สุดในทุกๆ ขนาดประชากร และเมื่อเปรียบเทียบประสิทธิภาพของค่า $RMSE$ ของ $\hat{N}_{GT}$ และ $\hat{N}_{NewGT}$ พบว่าค่า $RMSE$ ของ $\hat{N}_{GT}$ มีค่าต่ำกว่าของ $\hat{N}_{NewGT}$ ในทุกๆ สถานการณ์ และเมื่อเปรียบเทียบประสิทธิภาพของ $\hat{N}_{LB}$ และ $\hat{N}_{NewLB}$ พบว่าค่า $RMSE$ ของ $\hat{N}_{NewLB}$ มีค่าต่ำกว่าของ $\hat{N}_{LB}$ ในทุกๆ สถานการณ์

4.1.3 การจำลองภายใต้ตัวแบบการแจกแจงทวินามเชิงลบ

จากการจำลองภายใต้ตัวแบบการแจกแจงทวินามเชิงลบ ที่มีพารามิเตอร์ κ,π คือ $\kappa\in\{4,9\}$ และ $\pi\in\{0.50,0.80\}$ สำหรับการจำลองประชากรขนาดเท่ากับ $N=500, 1{,}000, 5{,}000$ $10{,}000, 50{,}000, 100{,}000$ ตามลำดับ รวมทั้งสิ้น 24 สถานการณ์ แสดงผลลัพธ์ได้ดังตารางที่ 4.6-4.9

4.1.3.1 พิจารณาค่าความเอนเอียงสัมพัทธ์

จากตาราง 4.6-4.9 เมื่อพารามิเตอร์ $\kappa=4,\pi=0.5$ พบว่า $RBias$ ของตัวประมาณขนาดประชากรทั้ง 4 มีค่าติดลบ เมื่อประชากรมีขนาดใหญ่ขึ้น และ $RBias$ ของ $\hat{N}_{GT}$ มีค่าต่ำลง เมื่อประชากรเพิ่มขึ้น หรือมีคุณสมบัติ asymptotic unbiased ในทางตรงกันข้าม พบว่าค่า $RBias$ ของ $\hat{N}_{NewLB}$ มีค่าสูงขึ้น เมื่อประชากรเพิ่มขึ้น และเมื่อเปรียบเทียบประสิทธิภาพของ

Ref. code: 25595809030157UGW

$\hat{N}_{GT}$ และ $\hat{N}_{NewGT}$ จากค่า ***RBias*** พบว่าค่า ***RBias*** ของ $\hat{N}_{NewGT}$ มีค่าต่ำกว่าของ $\hat{N}_{GT}$ ในทุกๆ ขนาดประชากร และเมื่อเปรียบเทียบประสิทธิภาพของ $\hat{N}_{LB}$ และ $\hat{N}_{NewLB}$ พบว่าค่า ***RBias*** ของ $\hat{N}_{LB}$ ให้ค่าต่ำกว่าของ $\hat{N}_{NewLB}$ ในทุกๆ ขนาดประชากร

พิจารณาพารามิเตอร์ $\kappa = 4, \pi = 0.8$ พบว่า ค่า ***RBias*** ของ $\hat{N}_{NewGT}$ มีค่าต่ำสุดเมื่อประชากรมีขนาดกลางและขนาดใหญ่ ($N = 5,000, 10,000, 50,000, 100,000$) และเมื่อเปรียบเทียบประสิทธิภาพของ $\hat{N}_{GT}$ และ $\hat{N}_{NewGT}$ จากค่า ***RBias*** พบว่าโดยภาพรวม $\hat{N}_{NewGT}$ ให้ค่า ***RBias*** ต่ำกว่า $\hat{N}_{GT}$ เกือบทุกสถานการณ์ และเมื่อเปรียบเทียบประสิทธิภาพของ $\hat{N}_{LB}$ และ $\hat{N}_{NewLB}$ พบว่าค่า ***RBias*** ของ $\hat{N}_{LB}$ ให้ค่าต่ำกว่าของ $\hat{N}_{NewLB}$ ในทุกๆ ขนาดประชากร

พิจารณาพารามิเตอร์ $\kappa = 9, \pi = 0.5$ พบว่า ค่า ***RBias*** ของ $\hat{N}_{NewGT}$ มีค่าต่ำสุดเมื่อประชากรมีขนาดกลางและขนาดใหญ่ และเมื่อเปรียบเทียบประสิทธิภาพของ $\hat{N}_{GT}$ และ $\hat{N}_{NewGT}$ จากค่า ***RBias*** พบว่า $\hat{N}_{NewGT}$ ให้ค่า ***RBias*** ต่ำกว่าของ $\hat{N}_{GT}$ เมื่อประชากรมีขนาดใหญ่ และเมื่อเปรียบเทียบประสิทธิภาพของ $\hat{N}_{LB}$ และ $\hat{N}_{NewLB}$ พบว่า $\hat{N}_{NewLB}$ ให้ค่า ***RBias*** ต่ำกว่าของ ***RBias*** เมื่อประชากรมีขนาดเล็ก

เมื่อพิจารณาพารามิเตอร์ $\kappa = 9, \pi = 0.8$ พบว่าค่า ***RBias*** ของ $\hat{N}_{NewGT}$ ต่ำสุดเกือบทุกสถานการณ์ นอกจากนี้ยังให้ค่า ***RBias*** ลดลงเมื่อประชากรเพิ่มขึ้นอีกด้วย และเมื่อเปรียบเทียบประสิทธิภาพระหว่าง $\hat{N}_{GT}$ และ $\hat{N}_{NewGT}$ พบว่า $\hat{N}_{NewGT}$ ให้ค่า ***RBias*** ต่ำกว่าของ $\hat{N}_{GT}$ เกือบทุกสถานการณ์ ยกเว้นกรณีประชากรมีขนาดเล็ก ($N = 500$) และเมื่อเปรียบเทียบประสิทธิภาพระหว่าง $\hat{N}_{LB}$ และ $\hat{N}_{NewLB}$ พบว่า $\hat{N}_{NewLB}$ ให้ค่า ***RBias*** ต่ำกว่าของ $\hat{N}_{LB}$ เมื่อประชากรมีขนาดเล็ก ($N = 500$)

4.1.3.2 พิจารณาค่าความแปรปรวนสัมพัทธ์

จากตาราง 4.6-4.9 เมื่อพิจารณาพารามิเตอร์ $\kappa = 4, \pi = 0.5, 0.8$ และ $\kappa = 9, \pi = 0.8$ ให้ข้อสรุปที่เหมือนกัน ซึ่งอธิบายได้ดังนี้ ค่า ***RVar*** ของ $\hat{N}_{NewLB}$ มีค่าต่ำที่สุดในทุกๆ สถานการณ์ และพบว่าตัวประมาณขนาดประชากรทั้ง 4 มีค่าเข้าใกล้ 0 เมื่อประชากรมีขนาดใหญ่ขึ้น และเมื่อเปรียบเทียบประสิทธิภาพของ $\hat{N}_{GT}$ และ $\hat{N}_{NewGT}$ พบว่า $\hat{N}_{GT}$ ให้ค่า ***RVar*** ต่ำกว่าของ ***RVar*** ในทุกๆ ขนาดประชากร แต่เมื่อพิจารณาประสิทธิภาพของ $\hat{N}_{LB}$ และ $\hat{N}_{NewLB}$ พบว่า $\hat{N}_{NewLB}$ ให้ค่า ***RVar*** ต่ำกว่าของ $\hat{N}_{LB}$ ในทุกๆ ขนาดประชากร

Ref. code: 25595809030157UGW

เมื่อพิจารณาพารามิเตอร์ $\kappa = 9, \pi = 0.5$ พบว่าค่า $RVar$ ของ $\hat{N}_{NewLB}$ มีค่าต่ำที่สุดในทุกๆ สถานการณ์ และพบว่าตัวประมาณขนาดประชากรทั้ง 4 มีค่าเข้าใกล้ 0 เมื่อประชากรมีขนาดใหญ่ขึ้น และเมื่อเปรียบเทียบประสิทธิภาพของ $\hat{N}_{GT}$ และ $\hat{N}_{NewGT}$ พบว่า $\hat{N}_{NewGT}$ ให้ค่า $RVar$ ต่ำกว่าของ $\hat{N}_{GT}$ เมื่อประชากรมีขนาดใหญ่ และเปรียบเทียบประสิทธิภาพของ $\hat{N}_{LB}$ และ $\hat{N}_{NewLB}$ พบว่า $\hat{N}_{NewLB}$ ให้ค่า $RVar$ ต่ำกว่าของ $\hat{N}_{LB}$ ในทุกๆ ขนาดประชากร

4.1.3.3 พิจารณาค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์

จากตาราง 4.6-4.9 เมื่อพิจารณาพารามิเตอร์ $\kappa = 4, \pi = 0.5$ พบว่า ค่า $RMSE$ ของ $\hat{N}_{GT}$ มีค่าต่ำสุดในทุกๆ สถานการณ์ และพบว่าตัวประมาณขนาดประชากรทั้ง 4 มีค่าเข้าใกล้ 0 เมื่อประชากรมีขนาดใหญ่ขึ้น และเมื่อเปรียบเทียบประสิทธิภาพของ $\hat{N}_{GT}$ และ $\hat{N}_{NewGT}$ พบว่า $\hat{N}_{GT}$ ให้ค่า $RMSE$ ต่ำกว่า $\hat{N}_{NewGT}$ ในทุกๆ ขนาดประชากร และเปรียบเทียบประสิทธิภาพของ $\hat{N}_{LB}$ และ $\hat{N}_{NewLB}$ พบว่า $\hat{N}_{NewLB}$ ให้ค่า $RMSE$ ต่ำกว่าของ $\hat{N}_{LB}$ เมื่อประชากรมีขนาดเล็ก ($N = 500, 1,000$)

เมื่อพารามิเตอร์ $\kappa = 4, \pi = 0.8$ พบว่า ค่า $RMSE$ ของ $\hat{N}_{NewLB}$ มีค่าต่ำ เมื่อประชากรมีขนาดเล็ก แต่เมื่อประชากรเพิ่มขึ้น พบว่า ค่า $RMSE$ ของ $\hat{N}_{GT}$ มีค่าต่ำสุด และเมื่อเปรียบเทียบประสิทธิภาพของ $\hat{N}_{GT}$ และ $\hat{N}_{NewGT}$ พบว่า $\hat{N}_{GT}$ ให้ค่า $RMSE$ ต่ำกว่า $\hat{N}_{NewGT}$ ในทุกๆ ขนาดประชากร และเปรียบเทียบประสิทธิภาพของ $\hat{N}_{LB}$ และ $\hat{N}_{NewLB}$ พบว่า $\hat{N}_{NewLB}$ ให้ค่า $RMSE$ ต่ำกว่าของ $\hat{N}_{LB}$ เมื่อประชากรมีขนาดเล็ก ($N = 500, 1,000$)

เมื่อพารามิเตอร์ $\kappa = 9, \pi = 0.5$ พบว่า ค่า $RMSE$ ของ $\hat{N}_{NewLB}$ มีค่าต่ำสุดเมื่อประชากรมีขนาดเล็ก ($N = 500, 1,000$) แต่พบว่าเมื่อประชากรเพิ่มขึ้น $\hat{N}_{GT}$ ให้ค่า $RMSE$ ต่ำที่สุด และเมื่อเปรียบเทียบประสิทธิภาพของ $\hat{N}_{GT}$ และ $\hat{N}_{NewGT}$ พบว่า $\hat{N}_{GT}$ ให้ค่า $RMSE$ ต่ำกว่าของ $\hat{N}_{NewGT}$ ทุกๆ ขนาดประชากร และเปรียบเทียบประสิทธิภาพของ $\hat{N}_{LB}$ และ $\hat{N}_{NewLB}$ พบว่า $\hat{N}_{NewLB}$ มีค่า $RMSE$ ต่ำกว่าของ $\hat{N}_{LB}$ เกือบทุกสถานการณ์ ยกเว้นกรณีประชากรมีขนาดใหญ่

เมื่อพารามิเตอร์ $\kappa = 9, \pi = 0.8$ พบว่า $RMSE$ ของ $\hat{N}_{NewLB}$ มีค่าต่ำสุด เมื่อประชากรมีขนาดเล็ก ($N = 500, 1,000$) แต่เมื่อประชากรมีขนาดใหญ่ขึ้น $\hat{N}_{GT}$ ให้ค่า $RMSE$ ต่ำที่สุด และเมื่อเปรียบเทียบประสิทธิภาพของ $\hat{N}_{GT}$ และ $\hat{N}_{NewGT}$ พบว่า $\hat{N}_{GT}$ ให้ค่า $RMSE$ ต่ำกว่าของ $\hat{N}_{NewGT}$ ทุกๆ ขนาดประชากร และเปรียบเทียบประสิทธิภาพของ $\hat{N}_{LB}$ และ $\hat{N}_{NewLB}$ พบว่า $\hat{N}_{NewLB}$ มีค่า $RMSE$ ต่ำกว่าของ $\hat{N}_{LB}$ เมื่อประชากรมีขนาดเล็ก ($N = 500, 1,000$)

Ref. code: 25595809030157UGW

ตารางที่ 4.1 แสดงค่าประมาณขนาดประชากร ค่าความเอนเอียงสัมพัทธ์ ค่าความแปรปรวนสัมพัทธ์และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ของตัวประมาณทั้ง 4 ตัว ภายใต้การแจกแจงปัวซง $p_i \sim Poi(\lambda)$ เมื่อกำหนดให้ $\lambda = 3$ และขนาดประชากร $N = 500, 1,000, 5,000, 10,000, 50,000, 100,000$

Estimator	$N = 500$				$N = 1,000$				$N = 5,000$			
	$\bar{\hat{N}}$	*RBias*	*RVar*	*RMSE*	$\bar{\hat{N}}$	*RBias*	*RVar*	*RMSE*	$\bar{\hat{N}}$	*RBias*	*RVar*	*RMSE*
GT	504.26	0.0085	0.0002*	0.0002*	1,006.15	0.0061	8.24E-05*	0.0001*	5,013.92	0.0028	1.52E-05*	2.29E-05
NewGT	506.07	0.0121	0.0003	0.0004	1,008.21	0.0082	0.0001	0.0002	5,016.61	0.0033	2.12E-05	3.22E-05
LB	502.80	0.0056	0.0009	0.0009	1,002.53	0.0025	0.0004	0.0004	5,003.19	0.0006	6.27E-05	6.31E-05
NewLB	501.71	0.0034*	0.0002*	0.0002*	1,001.19	0.0012*	0.0001	0.0001*	5,000.93	0.0002*	2.19E-05	2.20E-05*

Estimator	$N = 10,000$				$N = 50,000$				$N = 100,000$			
	$\bar{\hat{N}}$	*RBias*	*RVar*	*RMSE*	$\bar{\hat{N}}$	*RBias*	*RVar*	*RMSE*	$\bar{\hat{N}}$	*RBias*	*RVar*	*RMSE*
GT	10,018.98	0.0002	7.59E-06*	1.12E-05*	50,034.31	0.0007	1.55E-06*	2.11E-06*	100,055.90	0.0006	7.44E-07*	1.05E-06*
NewGT	10,024.22	0.0024	1.00E-05	1.59E-05	50,048.93	0.0010	1.86E-06	2.82E-06	100,069.30	0.0007	1.00E-06	1.48E-06
LB	10,004.63	0.0005	3.11E-05	3.13E-05	50,004.55	9.09E-05	6.49E-06	6.51E-06	100,008.10	8.14E-05	3.19E-06	3.19E-06
NewLB	10,001.02	0.0001*	1.12E-05	1.12E-05*	50,003.68	7.36E-05*	2.16E-06	2.16E-06	100,002.40	2.39E-05*	1.10E-06	1.10E-06

* หมายถึง *RBias, RVar* และ *RMSE* มีค่าต่ำที่สุด

Ref. code: 25595809030157UGW

ตารางที่ 4.2 แสดงค่าประมาณขนาดประชากร ค่าความเอนเอียงสัมพัทธ์ ค่าความแปรปรวนสัมพัทธ์และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ของตัวประมาณทั้ง 4 ตัว ภายใต้การแจกแจงปัวซง $p_i \sim Poi(\lambda)$ เมื่อกำหนดให้ $\lambda = 5$ และขนาดประชากร $N = 500, 1,000, 5,000, 10,000, 50,000, 100,000$

Estimator	$N=500$				$N=1,000$				$N=5,000$			
	$\bar{\hat{N}}$	*RBias*	*RVar*	*RMSE*	$\bar{\hat{N}}$	*RBias*	*RVar*	*RMSE*	$\bar{\hat{N}}$	*RBias*	*RVar*	*RMSE*
GT	500.91	0.0018	2.05E-05	2.38E-05	1,001.09	0.0011	8.83E-06*	1.00E-05	5,002.82	0.0006	1.60E-06*	1.92E-06
NewGT	501.28	0.0025	3.75E-05	4.40E-05	1,001.12	0.0011	1.00E-05	1.13E-05	5,002.21	0.0004	1.70E-06	1.89E-06
LB	501.60	0.0032	0.0001	0.0002	1,001.56	0.0016	4.30E-05	4.30E-05	5,001.22	0.0002	5.09E-06	5.15E-06
NewLB	500.28	0.0005*	1.94E-05*	1.97E-05*	1,000.30	0.0003*	9.56E-06	9.65E-06*	5,000.24	4.71E-05*	1.68E-06	1.69E-06*

Estimator	$N=10,000$				$N=50,000$				$N=100,000$			
	$\bar{\hat{N}}$	*RBias*	*RVar*	*RMSE*	$\bar{\hat{N}}$	*RBias*	*RVar*	*RMSE*	$\bar{\hat{N}}$	*RBias*	*RVar*	*RMSE*
GT	10,003.71	0.0004	8.44E-07*	9.81E-07	50,007.09	0.0001	1.48E-07*	1.68E-07*	100,009.30	9.28E-05	7.85E-08*	8.71E-08
NewGT	10,003.02	0.0003	9.29E-07	1.02E-06	50,005.84	0.0001	1.59E-07	1.72E-07	100,008.60	8.64E-05	8.19E-08	8.93E-08
LB	10,001.02	0.0001	2.39E-06	2.40E-06	50,000.83	1.67E-05	4.99E-07	4.99E-07	100,000.70	6.66E-06	2.30E-07	2.30E-07
NewLB	10,000.54	5.37E–05*	9.30E-07	9.33E-07*	50,000.34	6.83E-06*	1.86E-07	1.86E-07	100,000.20	1.68E-06*	8.59E-08	8.59E-08*

* หมายถึง *RBias, RVar* และ *RMSE* มีค่าต่ำที่สุด

Ref. code: 25595809030157UGW

ตารางที่ 4.3 แสดงค่าประมาณขนาดประชากร ค่าความเอนเอียงสัมพัทธ์ ค่าความแปรปรวนสัมพัทธ์และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ของตัวประมาณทั้ง 4 ตัว ภายใต้การแจกแจงปัวซง $p_i \sim Poi(\lambda)$ เมื่อกำหนดให้ $\lambda = 7$ และขนาดประชากร $N = 500, 1,000, 5,000, 10,000, 50,000, 100,000$

Estimator	$N = 500$				$N = 1,000$				$N = 5,000$			
	$\bar{\hat{N}}$	*RBias*	*RVar*	*RMSE*	$\bar{\hat{N}}$	*RBias*	*RVar*	*RMSE*	$\bar{\hat{N}}$	*RBias*	*RVar*	*RMSE*
GT	500.19	0.0004	2.39E-06*	2.55E-06*	1,000.26	0.0003	1.16E-06*	1.22E-06	5,000.52	0.0001	2.21E-07	2.32E-07
NewGT	500.36	0.0007	2.16E-05	2.22E-05	1,000.25	0.0002	1.20E-06	1.26E-06	5,000.40	8.05E-05	2.10E-07*	2.17E-07*
LB	501.20	0.0024	7.26E-05	7.84E-05	1,000.78	0.0008	9.58E-06	1.02E-05*	5,000.53	0.0001	6.10E-07	6.22E-07
NewLB	500.07	0.0001*	2.84E-06	2.86E-06	1,000.02	1.56E-05*	1.16E-06*	1.17E-06	5,000.10	1.93E-05*	2.25E-07	2.25E-07

Estimator	$N = 10,000$				$N = 50,000$				$N = 100,000$			
	$\bar{\hat{N}}$	*RBias*	*RVar*	*RMSE*	$\bar{\hat{N}}$	*RBias*	*RVar*	*RMSE*	$\bar{\hat{N}}$	*RBias*	*RVar*	*RMSE*
GT	10,000.77	7.67E-05	9.97E-08*	1.06E-07	50,001.43	2.87E-05	2.06E-08*	2.14E-08	100,002.20	2.15E-05	1.08E-08	1.13E-08
NewGT	10,000.46	4.65E-05	9.99E-08	1.02E-07	50,000.85	1.70E-05	2.13E-08	2.16E-08	100,000.70	7.27E-06	1.10E-08	1.11E-08
LB	10,000.78	7.79E-05	2.84E-07	2.90E-07	50,000.90	1.80E-05	4.76E-08	4.79E-08	100,000.80	8.02E-06	2.56E-08	2.57E-08
NewLB	10,000.00	4.84E-07*	1.00E-07	1.00E-07*	50,000.18	3.59E-06*	2.06E-08*	2.06E-08*	100,000.10	5.91E-07*	1.05E-08*	1.05E-08*

* หมายถึง *RBias*, *RVar* และ *RMSE* มีค่าต่ำที่สุด

Ref. code: 25595809030157UGW

ตารางที่ 4.4 แสดงค่าประมาณขนาดประชากร ค่าความเอนเอียงสัมพัทธ์ ค่าความแปรปรวนสัมพัทธ์และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ของตัวประมาณทั้ง 4 ตัว ภายใต้การแจกแจงปัวซงแบบผสมที่มีพารามิเตอร์ λ_1, λ_2 หรือ $p_i \sim 0.5Poi(\lambda_1) + 0.5Poi(\lambda_2)$ โดยที่ $(\lambda_1, \lambda_2) = (3,5)$ และขนาดประชากร $N = 500, 1,000, 5,000, 10,000, 50,000, 100,000$

Estimator	$N=500$				$N=1,000$				$N=5,000$			
	$\bar{\hat{N}}$	*RBias*	*RVar*	*RMSE*	$\bar{\hat{N}}$	*RBias*	*RVar*	*RMSE*	$\bar{\hat{N}}$	*RBias*	*RVar*	*RMSE*
GT	499.63	-0.0007*	0.0001*	0.0001*	999.20	-0.0008	4.88E-05*	4.94E-05*	5,000.91	0.0002*	1.04E-05*	1.05E-05*
NewGT	501.83	0.0037	0.0002	0.0002	1,001.74	0.0017	7.17E-05	7.48E-05	5,005.14	0.0010	1.48E-05	1.59E-05
LB	502.25	0.0045	0.0005	0.0005	1,003.10	0.0031	0.0002	0.0002	5,004.83	0.0010	3.37E-05	3.47E-05
NewLB	500.49	0.0010	0.0001*	0.0001*	1,000.23	0.0002*	5.25E-05	5.25E-05	4,994.90	-0.0010	1.05E-05	1.15E-05
Estimator	**$N=10,000$**				**$N=50,000$**				**$N=100,000$**			
	$\bar{\hat{N}}$	*RBias*	*RVar*	*RMSE*	$\bar{\hat{N}}$	*RBias*	*RVar*	*RMSE*	$\bar{\hat{N}}$	*RBias*	*RVar*	*RMSE*
GT	10,003.88	0.0004*	5.42E-06	5.57E-06*	50,033.18	0.0007	9.99E-07*	1.43E-06*	100,067.30	0.0007*	5.05E-07	9.58E-07*
NewGT	10,010.61	0.0011	7.21E-06	8.33E-06	50,060.96	0.0012	1.43E-06	2.92E-06	100,120.20	0.0012	7.21E-07	2.16E-06
LB	10,006.18	0.0006	1.54E-05	1.58E-05	50,023.84	0.0005*	3.15E-06	3.38E-06	100,038.10	0.0004	1.58E-06	1.73E-06
NewLB	9,989.69	-0.0010	5.22E-06*	6.28E-06	49,950.40	-0.0009	1.07E-06	2.05E-06	99,896.71	-0.0010	4.89E-07*	1.55E-06

* หมายถึง *RBias*, *RVar* และ *RMSE* มีค่าต่ำที่สุด

Ref. code: 25595809030157UGW

ตารางที่ 4.5 แสดงค่าประมาณขนาดประชากร ค่าความเอนเอียงสัมพัทธ์ ค่าความแปรปรวนสัมพัทธ์และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ของตัวประมาณทั้ง 4 ตัว ภายใต้การแจกแจงปัวซงแบบผสมที่มีพารามิเตอร์ λ_1, λ_2 หรือ $p_i \sim 0.5Poi(\lambda_1) + 0.5Poi(\lambda_2)$ โดยที่ $(\lambda_1, \lambda_2) = (7,5)$ และขนาดประชากร $N = 500, 1,000, 5,000, 10,000, 50,000, 100,000$

Estimator	$N = 500$				$N = 1,000$				$N = 5,000$			
	$\bar{\hat{N}}$	*RBias*	*RVar*	*RMSE*	$\bar{\hat{N}}$	*RBias*	*RVar*	*RMSE*	$\bar{\hat{N}}$	*RBias*	*RVar*	*RMSE*
GT	500.40	0.0008*	0.0001*	0.0001*	1,002.27	0.0022	5.63E-05	6.14E-05	5,021.19	0.0042	1.07E-05	2.87E-05
NewGT	503.84	0.0077	0.0005	0.0005	1,007.11	0.0071	0.0001	0.0002	5,037.47	0.0075	2.51E-05	8.13E-05
LB	502.353	0.0047	0.0005	0.0006	1,003.13	0.0031	0.0002	0.0002	5,007.44	0.0015	3.20E-05	3.42E-05
NewLB	501.262	0.0025	0.0001*	0.0001*	1,001.60	0.0016*	5.37E-05*	5.62E-05*	5,000.07	0.0002*	9.60E-06*	9.64E-06*

Estimator	$N = 10,000$				$N = 50,000$				$N = 100,000$			
	$\bar{\hat{N}}$	*RBias*	*RVar*	*RMSE*	$\bar{\hat{N}}$	*RBias*	*RVar*	*RMSE*	$\bar{\hat{N}}$	*RBias*	*RVar*	*RMSE*
GT	10,046.91	0.0047	5.60E-06	2.76E-05	50,257.97	0.0051	1.11E-06	2.77E-05	100,521.20	0.0052	5.19E-07*	2.76E-05
NewGT	10,082.97	0.0083	1.71E-05	8.59E-05	50,441.89	0.0088	6.36E-06	8.44E-05	100,879.50	0.0087	3.71E-06	8.10E-05
LB	10,012.49	0.0012	1.58E-05	1.74E-05	50,046.94	0.0009	3.27E-06	4.15E-06	100,085.00	0.0008	1.55E-06	2.27E-06
NewLB	10,003.32	0.0002*	4.23E-06*	4.34E-06*	50,015.99	0.0002*	9.60E-07*	1.06E-06*	100,032.70	0.0002*	5.61E-07	6.68E-07*

* หมายถึง *RBias, RVar* และ *RMSE* มีค่าต่ำที่สุด

Ref. code: 25595809030157UGW

ตารางที่ 4.6 แสดงค่าประมาณขนาดประชากร ค่าความเอนเอียงสัมพัทธ์ ค่าความแปรปรวนสัมพัทธ์และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ของตัวประมาณทั้ง 4 ตัว ภายใต้การแจกแจงทวินามเชิงลบที่มีพารามิเตอร์ π และ κ , $p_i \sim NB(\kappa,\pi)$ โดยที่ $\kappa = 4,\ \pi = 0.5$ และขนาดประชากร $N = 500, 1,000, 5,000, 10,000,$ $50,000, 100,000$

Estimator	$N = 500$				$N = 1,000$				$N = 5,000$			
	$\bar{\hat{N}}$	*RBias*	*RVar*	*RMSE*	$\bar{\hat{N}}$	*RBias*	*RVar*	*RMSE*	$\bar{\hat{N}}$	*RBias*	*RVar*	*RMSE*
GT	495.16	-0.0097	0.0004	0.0004*	992.68	-0.0073	0.0002	0.0002*	4,986.30	-0.0027	3.80E-05	4.55E-05*
NewGT	500.21	0.0004*	0.0018	0.0018	999.36	-0.0006*	0.0005	0.0005	4,998.77	-0.0002*	9.31E-05	9.32E-05
LB	502.91	0.0058	0.0016	0.0016	1,001.53	0.0015	0.0007	0.0007	4,991.53	-0.0017	0.0001	0.0001
NewLB	495.33	-0.0093	0.0003*	0.0004	988.92	-0.0110	0.0001*	0.0003	4,937.46	-0.0125	3.17E-05*	0.0002

Estimator	$N = 10,000$				$N = 50,000$				$N = 100,000$			
	$\bar{\hat{N}}$	*RBias*	*RVar*	*RMSE*	$\bar{\hat{N}}$	*RBias*	*RVar*	*RMSE*	$\bar{\hat{N}}$	*RBias*	*RVar*	*RMSE*
GT	9,983.09	-0.0017	1.91E-05	2.19E-05*	49,976.78	-0.0005	3.78E-06	4.00E-06*	99,973.76	-0.0003	1.87E-06	1.93E-06*
NewGT	9,998.86	-0.0001*	4.69E-05	4.69E-05	49,997.74	-4.51E-05*	9.70E-06	9.70E-06	99,996.34	-3.66E-05*	4.87E-06	4.87E-06
LB	9,978.98	-0.0021	5.50E-05	5.94E-05	49,879.98	-0.0024	1.11E-05	1.68E-05	99,753.55	-0.0025	5.48E-06	1.16E-05
NewLB	9,871.74	-0.0128	1.61E-05*	0.0002	49,318.56	-0.0136	3.74E-06*	0.0002	98,596.97	-0.0140	1.84E-06*	0.0002

* หมายถึง *RBias*, *RVar* และ *RMSE* มีค่าต่ำที่สุด

Ref. code: 25595809030157UGW

ตารางที่ 4.7 แสดงค่าประมาณขนาดประชากร ค่าความเอนเอียงสัมพัทธ์ ค่าความแปรปรวนสัมพัทธ์และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ของตัวประมาณทั้ง 4 ตัว ภายใต้การแจกแจงทวินามเชิงลบที่มีพารามิเตอร์ π และ κ , $p_i \sim NB(\kappa, \pi)$ โดยที่ $\kappa = 4,\ \pi = 0.8$ และขนาดประชากร $N = 500, 1,000, 5,000, 10,000,$ $50,000,\ 100,000$

Estimator	$N = 500$				$N = 1,000$				$N = 5,000$			
	$\bar{\hat{N}}$	*RBias*	*RVar*	*RMSE*	$\bar{\hat{N}}$	*RBias*	*RVar*	*RMSE*	$\bar{\hat{N}}$	*RBias*	*RVar*	*RMSE*
GT	508.39	0.0167*	0.0281	0.0284	982.71	-0.0172	0.0096	0.0099	4,932.02	-0.0135	0.0019	0.0021*
NewGT	552.70	0.1054	0.1593	0.1705	1,031.08	0.0311	0.0264	0.0274	5,034.28	0.0068*	0.0058	0.0058
LB	509.33	0.0186	0.0349	0.0352	1,001.45	0.0014*	0.0166	0.0166	4,945.39	-0.0109	0.0030	0.0031
NewLB	462.01	-0.0759	0.0049*	0.0107*	922.75	-0.0772	0.0024*	0.0083*	4,598.62	-0.0803	0.0004*	0.0069

Estimator	$N = 10,000$				$N = 50,000$				$N = 100,000$			
	$\bar{\hat{N}}$	*RBias*	*RVar*	*RMSE*	$\bar{\hat{N}}$	*RBias*	*RVar*	*RMSE*	$\bar{\hat{N}}$	*RBias*	*RVar*	*RMSE*
GT	9,904.50	-0.0095	0.0009	0.0010*	49,856.47	-0.0028	0.0002	0.0002*	99,827.77	-0.0017	9.84E-05	0.0001*
NewGT	10,043.28	0.0043*	0.0035	0.0036	50,125.57	0.0025*	0.0008	0.0009	100,122.30	0.0012*	0.0004	0.0004
LB	9,860.35	-0.0139	0.0015	0.0017	49,222.25	-0.0155	0.0003	0.0005	98,396.17	-0.0160	0.0001	0.0004
NewLB	9,185.16	-0.0815	0.0002*	0.0068	45,912.71	-0.0187	4.65E-05*	0.0067	91,809.66	-0.0819	2.38E-05*	0.0067

* หมายถึง *RBias*, *RVar* และ *RMSE* มีค่าต่ำที่สุด

Ref. code: 25595809030157UGW

ตารางที่ 4.8 แสดงค่าประมาณขนาดประชากร ค่าความเอนเอียงสัมพัทธ์ ค่าความแปรปรวนสัมพัทธ์และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ของตัวประมาณทั้ง 4 ตัว ภายใต้การแจกแจงทวินามเชิงลบที่มีพารามิเตอร์ π และ κ , $p_i \sim NB(\kappa,\pi)$ โดยที่ $\kappa = 9,\ \pi = 0.5$ และขนาดประชากร $N = 500, 1,000, 5,000, 10,000, 50,000, 100,000$

Estimator	$N=500$				$N=1,000$				$N=5,000$			
	$\bar{\hat{N}}$	*RBias*	*RVar*	*RMSE*	$\bar{\hat{N}}$	*RBias*	*RVar*	*RMSE*	$\bar{\hat{N}}$	*RBias*	*RVar*	*RMSE*
GT	500.09	0.0002*	1.36E-05	1.36E-05	999.91	-9.02E-05	3.00E-06	3.01E-06	4,999.44	-0.0001	4.62E-07	4.75E-07*
NewGT	504.38	0.0087	0.0085	0.0086	1,000.85	0.0008	4.81E-05	4.88E-05	5,000.02	3.39E-06*	5.19E-07	5.19E-07
LB	503.01	0.0060	0.0006	0.0007	1,002.19	0.0022	8.10E-05	8.58E-05	5,001.19	0.0002	2.04E-06	2.10E-06
NewLB	500.15	0.0003	9.55E-06*	9.64E-06*	999.92	-7.99E-05*	2.74E-06*	2.74E-06*	4,998.46	-0.0003	4.36E-07*	5.31E-07

Estimator	$N=10,000$				$N=50,000$				$N=100,000$			
	$\bar{\hat{N}}$	*RBias*	*RVar*	*RMSE*	$\bar{\hat{N}}$	*RBias*	*RVar*	*RMSE*	$\bar{\hat{N}}$	*RBias*	*RVar*	*RMSE*
GT	9,999.43	-5.71E-05	2.14E-07	2.18E-07*	49,998.57	-2.85E-05	5.04E-08	5.12E-08*	99,998.07	-1.92E-05	2.66E-08	2.70E-08*
NewGT	10,000.18	1.76E-05*	2.37E-07	2.38E-07	49,999.47	-1.05E-05*	5.36E-08	5.37E-08	99,998.88	-1.11E-05*	2.82E-08	2.83E-08
LB	10,000.89	8.89E-05	7.74E-07	7.82E-07	49,999.09	-1.81E-05	1.47E-07	1.47E-07	99,997.55	-2.45E-05	7.29E-08	7.35E-08
NewLB	9,996.84	-0.0003	1.93E-07*	2.92E-07	49,983.11	-0.0003	4.73E-08*	1.61E-07	99,966.61	-0.0003	2.30E-08*	1.34E-07

* หมายถึง *RBias*, *RVar* และ *RMSE* มีค่าต่ำที่สุด

Ref. code: 25595809030157UGW

ตารางที่ 4.9 แสดงค่าประมาณขนาดประชากร ค่าความเอนเอียงสัมพัทธ์ ค่าความแปรปรวนสัมพัทธ์และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ของตัวประมาณทั้ง 4 ตัว ภายใต้การแจกแจงทวินามเชิงลบที่มีพารามิเตอร์ π และ κ , $p_i \sim NB(\kappa, \pi)$ โดยที่ $\kappa = 9,\ \pi = 0.8$ และขนาดประชากร $N = 500, 1,000, 5,000, 10,000,$ $50,000,\ 100,000$

Estimator	$N = 500$				$N = 1,000$				$N = 5,000$			
	$\overline{\hat{N}}$	*RBias*	*RVar*	*RMSE*	$\overline{\hat{N}}$	*RBias*	*RVar*	*RMSE*	$\overline{\hat{N}}$	*RBias*	*RVar*	*RMSE*
GT	497.76	-0.0044*	0.0010	0.0010	993.12	-0.0069	0.0005	0.0006*	4,982.67	-0.0035	0.0001	0.0001*
NewGT	505.05	0.0101	0.0020	0.0022	1,001.68	0.0017*	0.0010	0.0010	4,993.99	-0.0012	0.0002	0.0002
LB	506.33	0.0126	0.0048	0.0050	1,003.56	0.0036	0.0019	0.0019	4,997.06	-0.0006*	0.0003	0.0003
NewLB	495.21	-0.0096	0.0008*	0.0009*	987.39	-0.0126	0.0004*	0.0006*	4,934.47	-0.0131	8.85E-05*	0.0003
Estimator	$N = 10,000$				$N = 50,000$				$N = 100,000$			
	$\overline{\hat{N}}$	*RBias*	*RVar*	*RMSE*	$\overline{\hat{N}}$	*RBias*	*RVar*	*RMSE*	$\overline{\hat{N}}$	*RBias*	*RVar*	*RMSE*
GT	9,983.07	-0.0017	5.59E-05	5.88E-05*	49,974.36	-0.0005	1.04E-05	1.07E-05*	99,968.03	-0.0003	5.12E-06	5.22E-06*
NewGT	9,996.53	-0.0003*	0.0001	0.0001	49,991.07	-0.0002*	2.23E-05	2.23E-05	99,993.48	-6.52E-05*	1.17E-05	1.17E-05
LB	9,994.15	-0.0006	0.0001	0.0001	49,942.90	-0.0011	2.88E-05	3.02E-05	99,873.90	-0.0013	1.48E-05	1.64E-05
NewLB	9,869.66	-0.0130	4.54E-05*	0.0002	49,332.98	-0.0133	8.57E-06*	0.0002	98,658.34	-0.0134	4.19E-06*	0.0002

* หมายถึง *RBias*, *RVar* และ *RMSE* มีค่าต่ำที่สุด

Ref. code: 25595809030157UGW

4.2 เปรียบเทียบประสิทธิภาพของช่วงเชื่อมั่นของขนาดประชากร

ผลจากการจำลองสถานการณ์ได้แสดงไว้ดังตารางที่ 4.10-4.18 ซึ่งแบ่งออกเป็น 3 ส่วน คือ ผลจากการจำลองภายใต้ตัวแบบการแจกแจงปัวซงที่มีพารามิเตอร์ λ การแจกแจงปัวซงแบบผสมที่มีพารามิเตอร์ λ_1,λ_2 และการแจกแจงทวินามเชิงลบที่มีพารามิเตอร์ κ,π โดยพิจารณาจากค่าประมาณสัมประสิทธิ์ความเชื่อมั่นของช่วงเชื่อมั่นของขนาดประชากรและค่าประมาณความกว้างเฉลี่ยของช่วงเชื่อมั่นโดยตัวประมาณทั้ง 4 โดยผลการศึกษาแสดงได้ดังนี้

4.2.1 การจำลองภายใต้ตัวแบบการแจกแจงปัวซงที่มีพารามิเตอร์ λ

จากตารางที่ 4.10 แสดงค่าประมาณสัมประสิทธิ์ความเชื่อมั่นและค่าประมาณความกว้างเฉลี่ยของช่วงเชื่อมั่นของตัวประมาณทั้ง 4 ภายใต้การแจกแจงปัวซงที่มีพารามิเตอร์ λ เมื่อกำหนดพารามิเตอร์ $\lambda = 3$ และขนาดประชากร $N = 500, 1{,}000, 5{,}000, 10{,}000, 50{,}000, 100{,}000$ พบว่า ช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{NewLB}$ ให้ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นสูงกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนดในทุกๆ ขนาดประชากรหรือให้ค่าประมาณสูงกว่า 0.9464 นอกจากนี้พบว่า เมื่อประชากร $N = 50{,}000$ ช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{LB}$ ให้ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นสูงกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนดอีกด้วย โดยช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{NewLB}$ ให้ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นสูงที่สุดและให้ค่าประมาณความกว้างเฉลี่ยของช่วงเชื่อมั่นที่กว้างที่มีความกว้างเป็นอันดับสองในทุกๆ ขนาดประชากร และส่วนใหญ่ถึงแม้ว่าช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{GT}$ ให้ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนดในทุกๆ ขนาดประชากร แต่พบว่าให้ค่าประมาณความกว้างเฉลี่ยของช่วงเชื่อมั่นที่แคบที่สุด

จากตารางที่ 4.11 แสดงค่าประมาณสัมประสิทธิ์ความเชื่อมั่นและค่าประมาณความกว้างเฉลี่ยของช่วงเชื่อมั่นของตัวประมาณทั้ง 4 ภายใต้การแจกแจงปัวซงที่มีพารามิเตอร์ λ เมื่อกำหนดพารามิเตอร์ $\lambda = 5$ และขนาดประชากร $N = 500, 1{,}000, 5{,}000, 10{,}000, 50{,}000, 100{,}000$ พบว่าช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{NewLB}$ ให้ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นสูงกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนดในทุกๆ ขนาดประชากร และยังพบว่า เมื่อ $N = 500{,}1{,}000$ ยังมีช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{GT}$ ที่ให้ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นสูงกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด และพบว่า ช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{NewLB}$ ให้ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นสูงกว่าของ $\hat{N}_{GT}$ แต่พบว่า ช่วงเชื่อมั่นของขนาดประชากรโดย

$\hat{N}_{GT}$ ให้ค่าประมาณความกว้างเฉลี่ยของช่วงเชื่อมั่นต่ำกว่า และพบว่าเมื่อ $N = 5{,}000$ ยังมีช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{NewGT}$ ที่ให้ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นสูงกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด และพบว่า ช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{NewLB}$ ให้ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นสูงกว่าของ $\hat{N}_{NewGT}$ แต่พบว่า ช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{NewGT}$ ให้ค่าประมาณความกว้างเฉลี่ยของช่วงเชื่อมั่นต่ำกว่า และยังพบว่า เมื่อ $N = 10{,}000$ และ $N = 100{,}000$ ยังมีช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{LB}$ ที่ให้ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นสูงกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด และพบว่า ช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{NewLB}$ ให้ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นสูงกว่าของ $\hat{N}_{LB}$ และยังพบว่า ช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{NewLB}$ ให้ค่าประมาณความกว้างเฉลี่ยของช่วงเชื่อมั่นแคบกว่าอีกด้วย

จากตารางที่ 4.12 แสดงค่าประมาณสัมประสิทธิ์ความเชื่อมั่นและค่าประมาณความกว้างเฉลี่ยของช่วงเชื่อมั่นของตัวประมาณทั้ง 4 ภายใต้การแจกแจงปัวซงที่มีพารามิเตอร์ λ เมื่อพารามิเตอร์ $\lambda = 7$ และขนาดประชากร $N = 500, 1{,}000, 5{,}000, 10{,}000, 50{,}000, 100{,}000$ เมื่อ $N = 500$ พบว่า ช่วงเชื่อมั่นของขนาดประชากรโดยตัวประมาณขนาดประชากรทั้ง 4 ให้ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นต่ำกว่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด ถึงแม้ว่า ช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{NewLB}$ ให้ค่าประมาณช่วงเชื่อมั่นต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด แต่ให้ค่าประมาณความกว้างเฉลี่ยของช่วงเชื่อมั่นที่แคบที่สุด เมื่อ $N = 1{,}000$ พบว่า ช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{GT}, \hat{N}_{NewGT}$ ให้ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นสูงกว่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด และถึงแม้ว่า ช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{NewLB}$ ให้ค่าประมาณช่วงเชื่อมั่นต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด แต่ให้ค่าประมาณความกว้างเฉลี่ยของช่วงเชื่อมั่นที่แคบที่สุด และเมื่อ $N = 5{,}000, 10{,}000$ และ $N = 50{,}000$ พบว่า ช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{GT}, \hat{N}_{NewGT}, \hat{N}_{NewLB}$ ให้ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นสูงกว่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด ถึงแม้ว่า ช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{NewLB}$ ให้ค่าประมาณช่วงเชื่อมั่นต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด แต่ให้ค่าประมาณความกว้างเฉลี่ยของช่วงเชื่อมั่นที่แคบที่สุด เมื่อ $N = 100{,}000$ พบว่า ช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{NewGT}, \hat{N}_{LB}$ ให้ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นสูงกว่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด และยังพบว่า ช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{NewGT}$ ให้ค่าประมาณความกว้างเฉลี่ยของช่วงเชื่อมั่นที่แคบที่สุด

4.2.2 การจำลองภายใต้ตัวแบบการแจกแจงปัวซงแบบผสมที่มีพารามิเตอร์ λ_1, λ_2

จากตารางที่ 4.13 แสดงค่าประมาณสัมประสิทธิ์ความเชื่อมั่นและค่าประมาณความกว้างเฉลี่ยของช่วงเชื่อมั่นของตัวประมาณทั้ง 4 ภายใต้การแจกแจงปัวซงแบบผสมที่มีพารามิเตอร์ λ_1, λ_2 หรือ $p_i \sim 0.5Poi(\lambda_1) + 0.5Poi(\lambda_2)$ โดยที่ $(\lambda_1, \lambda_2) = (3,5)$ และ $N = 500,\ 1,000,\ 5,000,$ $10,000,\ 50,000,\ 100,000$

เมื่อพิจารณาที่ประชากรขนาดเล็ก $N = 500, 1,000$ พบว่า ช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{NewLB}$ ให้ค่าสัมประสิทธิ์ความเชื่อมั่นสูงกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด และช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{LB}$ ให้ค่าสัมประสิทธิ์ความเชื่อมั่นสูงเป็นอันดับสอง และพบว่า ช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{LB}$ ให้ค่าประมาณความกว้างเฉลี่ยของช่วงเชื่อมั่นที่กว้างที่สุด ช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{NewGT}$ ให้ค่าสัมประสิทธิ์ความเชื่อมั่นต่ำที่สุด และพบว่า ช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{NewGT}$ ให้ค่าประมาณความกว้างเฉลี่ยของช่วงเชื่อมั่นที่แคบเป็นอันดับสอง ช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{GT}$ ให้ค่าสัมประสิทธิ์ความเชื่อมั่นต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด แต่พบว่า ช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{GT}$ ให้ค่าประมาณความกว้างเฉลี่ยของช่วงเชื่อมั่นที่แคบที่สุด

เมื่อพิจารณาประชากร $N = 5,000$ พบว่า ช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{LB}, \hat{N}_{NewLB}$ ให้ค่าสัมประสิทธิ์ความเชื่อมั่นสูงกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด โดยช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{LB}$ ให้ค่าสัมประสิทธิ์ความเชื่อมั่นสูงที่สุด และพบว่า ช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{LB}$ ให้ค่าประมาณความกว้างเฉลี่ยของช่วงเชื่อมั่นที่กว้างที่สุด และแม้ว่า ช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{GT}$ ให้ค่าสัมประสิทธิ์ความเชื่อมั่นต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด แต่พบว่า ช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{GT}$ ให้ค่าประมาณความกว้างเฉลี่ยของช่วงเชื่อมั่นที่แคบที่สุด

เมื่อพิจารณาประชากร $N = 10,000, 50,000$ พบว่า ช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{LB}$ ให้ค่าสัมประสิทธิ์ความเชื่อมั่นสูงกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด และพบว่า ช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{LB}$ ให้ค่าประมาณความกว้างเฉลี่ยของช่วงเชื่อมั่นที่กว้างที่สุด และถึงแม้ว่า ช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{GT}$ ให้ค่าสัมประสิทธิ์ความเชื่อมั่นต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด แต่พบว่า ช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{GT}$ ให้ค่าประมาณความกว้างเฉลี่ยของช่วงเชื่อมั่นที่แคบที่สุด

จากตารางที่ 4.14 แสดงค่าประมาณสัมประสิทธิ์ความเชื่อมั่นและค่าประมาณความกว้างเฉลี่ยของช่วงเชื่อมั่นของตัวประมาณทั้ง 4 ภายใต้การแจกแจงปัวซงแบบผสมที่มีพารามิเตอร์ λ_1, λ_2

Ref. code: 25595809030157UGW

หรือ $p_i \sim 0.5Poi(\lambda_1)+0.5Poi(\lambda_2)$ โดยที่ $(\lambda_1,\lambda_2)=(7,5)$ และ $N=500,\ 1{,}000,\ 5{,}000,$ $10{,}000,\ 50{,}000,\ 100{,}000$ เมื่อพิจารณาที่ประชากรขนาดเล็ก $N=500,\ 1{,}000$ พบว่า ช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{NewLB}$ ให้ค่าสัมประสิทธิ์ความเชื่อมั่นสูงกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด และถึงแม้ว่าช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{GT}$ ให้ค่าสัมประสิทธิ์ความเชื่อมั่นต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด แต่พบว่า ช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{GT}$ ให้ค่าประมาณความกว้างเฉลี่ยของช่วงเชื่อมั่นที่แคบที่สุด

ช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{GT}$ ให้ค่าสัมประสิทธิ์ความเชื่อมั่นต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด แต่พบว่า ช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{GT}$ ให้ค่าประมาณความกว้างเฉลี่ยของช่วงเชื่อมั่นที่แคบที่สุด

ช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{NewGT}$ ให้ค่าสัมประสิทธิ์ความเชื่อมั่นต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด แต่พบว่า ช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{GT}$ ให้ค่าประมาณความกว้างเฉลี่ยของช่วงเชื่อมั่นที่แคบเป็นอันดับสอง

เมื่อพิจารณาประชากรขนาด $N=50{,}000$ พบว่า ช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{LB}$ ให้ค่าสัมประสิทธิ์ความเชื่อมั่นสูงกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด แต่พบว่า ช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{LB}$ ให้ค่าประมาณความกว้างเฉลี่ยของช่วงเชื่อมั่นที่กว้างที่สุด

ช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{GT}$ ให้ค่าสัมประสิทธิ์ความเชื่อมั่นต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด แต่พบว่า ช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{GT}$ ให้ค่าประมาณความกว้างเฉลี่ยของช่วงเชื่อมั่นที่แคบที่สุด

เมื่อพิจารณาประชากร $N=100{,}000$ พบว่า ช่วงเชื่อมั่นของขนาดประชากรทั้ง 4 ให้ค่าสัมประสิทธิ์ความเชื่อมั่นต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด โดยช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{GT}, \hat{N}_{LB}$ ให้ค่าสัมประสิทธิ์ความเชื่อมั่นสูงที่สุด และพบว่า ช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{GT}$ ให้ค่าประมาณความกว้างเฉลี่ยของช่วงเชื่อมั่นที่แคบที่สุด รองลงมาคือ ช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{NewGT}$, $\hat{N}_{NewLB}$ และ $\hat{N}_{LB}$ ตามลำดับ

Ref. code: 25595809030157UGW

4.2.3 การจำลองภายใต้ตัวแบบการแจกแจงทวินามเชิงลบ ที่มีพารามิเตอร์ κ,π

จากตารางที่ 4.15 แสดงค่าประมาณสัมประสิทธิ์ความเชื่อมั่นและค่าประมาณความกว้างเฉลี่ยของช่วงเชื่อมั่นของตัวประมาณทั้ง 4 ภายใต้การแจกแจงทวินามเชิงลบ (Negative Binomial distribution) ที่มีพารามิเตอร์ κ,π , $p_i \sim NB(\kappa,\pi)$ เมื่อพิจารณาพารามิเตอร์ $\kappa=4,\pi=0.5$ สำหรับขนาดประชากร $N=500,1{,}000, 5{,}000 ,10{,}000 ,50{,}000 ,100{,}000$ พบว่า ช่วงเชื่อมั่นของขนาดประชากรโดยตัวประมาณทั้ง 4 ให้ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนดในทุกขนาดประชากร และพบว่า ช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{LB}$ ให้ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นสูงที่สุดในทุกๆ ขนาดประชากร ยกเว้นกรณีที่ $N=100{,}000$ ถึงแม้ว่า ช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{LB}$ ให้ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นสูงที่สุด แต่พบว่าช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{LB}$ ให้ค่าประมาณความกว้างเฉลี่ยของช่วงเชื่อมั่นกว้างที่สุดในทุกๆ ขนาดประชากร นอกจากนี้พบว่า ช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{NewGT}$ ให้ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นต่ำเกือบทุกสถานการณ์ แต่พบว่า ช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{NewGT}$ ให้ค่าประมาณความกว้างเฉลี่ยของช่วงเชื่อมั่นที่แคบที่สุดเกือบทุกสถานการณ์เช่นกัน

จากตารางที่ 4.16 แสดงค่าประมาณสัมประสิทธิ์ความเชื่อมั่นและค่าประมาณความกว้างเฉลี่ยของช่วงเชื่อมั่นของตัวประมาณทั้ง 4 ภายใต้การแจกแจงทวินามเชิงลบ (Negative Binomial distribution) ที่พารามิเตอร์ κ,π , $p_i \sim NB(\kappa,\pi)$ พิจารณาพารามิเตอร์ $\kappa=4,\pi=0.8$ สำหรับขนาดประชากร $N=500,1{,}000, 5{,}000 ,10{,}000 ,50{,}000 ,100{,}000$ พบว่า ช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{NewLB}$ ให้ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นสูงกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนดเมื่อประชากรขนาด $N=500,1{,}000,5{,}000$ แต่ก็ยังพบว่าช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{NewLB}$ ให้ค่าประมาณความกว้างเฉลี่ยของช่วงเชื่อมั่นกว้างที่สุดในทุกๆ สถานการณ์ นอกจากนี้พบว่า ช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{NewGT}$ ให้ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นต่ำกว่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด แต่ช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{NewGT}$ ให้ค่าประมาณความกว้างเฉลี่ยของช่วงเชื่อมั่นที่แคบที่สุดในทุกๆ สถานการณ์

จากตารางที่ 4.17 แสดงค่าประมาณสัมประสิทธิ์ความเชื่อมั่นและค่าประมาณความกว้างเฉลี่ยของช่วงเชื่อมั่นของตัวประมาณทั้ง 4 ภายใต้การแจกแจงทวินามเชิงลบ (Negative Binomial distribution) ที่มีพารามิเตอร์ κ,π , $p_i \sim NB(\kappa,\pi)$ พิจารณาพารามิเตอร์ $\kappa=9,\pi=0.5$ สำหรับขนาดประชากร $N=500,1{,}000, 5{,}000 ,10{,}000 ,50{,}000 ,100{,}000$ พบว่า ช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{GT}, \hat{N}_{NewGT}$ ให้ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นสูงกว่าค่า

Ref. code: 25595809030157UGW

สัมประสิทธิ์ความเชื่อมั่นที่กำหนดเมื่อประชากรขนาด $N = 10,000$ และพบว่า ช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{GT}$ ให้ค่าประมาณความกว้างเฉลี่ยของช่วงเชื่อมั่นแคบที่สุดเมื่อประชากรมีขนาดเล็ก

และเมื่อประชากรมีขนาดใหญ่ พบว่าช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{NewLB}$ ให้ค่าประมาณความกว้างเฉลี่ยของช่วงเชื่อมั่นแคบที่สุด ถึงแม้ว่าช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{NewLB}$ จะให้ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นต่ำกว่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนดก็ตาม นอกจากนี้ยังพบว่า ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นของตัวประมาณขนาดประชากรทั้ง 4 มีค่าสูงและใกล้เคียงกับค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนดในทุกๆ ขนาดประชากร

จากตารางที่ 4.18 แสดงค่าประมาณสัมประสิทธิ์ความเชื่อมั่นและค่าประมาณความกว้างเฉลี่ยของช่วงเชื่อมั่นของตัวประมาณทั้ง 4 ภายใต้การแจกแจงทวินามเชิงลบ (Negative Binomial distribution) ที่มีพารามิเตอร์ κ, π , $p_i \sim NB(\kappa, \pi)$ พิจารณาพารามิเตอร์ $\kappa = 9, \pi = 0.8$ สำหรับขนาดประชากร $N = 500, 1,000, 5,000, 10,000, 50,000, 100,000$ พบว่าช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{NewLB}$ ให้ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นสูงกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนดเมื่อประชากรขนาด $N = 500, 1,000, 5,000$ แต่ช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{NewLB}$ ก็ให้ค่าประมาณความกว้างเฉลี่ยของช่วงเชื่อมั่นที่กว้างที่สุดในทุกๆ ขนาดประชากร และถึงแม้ว่าช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{NewGT}$ ให้ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นต่ำกว่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด แต่ก็ให้ค่าประมาณความกว้างเฉลี่ยของช่วงเชื่อมั่นที่แคบที่สุดในทุกๆ ขนาดประชากรเช่นเดียวกัน

Ref. code: 25595809030157UGW

ตารางที่ 4.10 แสดงค่าประมาณสัมประสิทธิ์ความเชื่อมั่นและค่าประมาณความกว้างเฉลี่ยของช่วงเชื่อมั่นของตัวประมาณทั้ง 4 ภายใต้การแจกแจงปัวซง $p_i \sim Poi(\lambda)$ เมื่อกำหนดให้ $\lambda = 3$ และขนาดประชากร $N = 500, 1,000, 5,000, 10,000, 50,000, 100,000$

Estimator	$N = 500$		$N = 1,000$		$N = 5,000$	
	$1-\hat{\alpha}$	Average Length	$1-\hat{\alpha}$	Average Length	$1-\hat{\alpha}$	Average Length
GT	0.907	26.062	0.886	34.622	0.886	74.831
NewGT	0.817	25.997	0.818	35.414	0.828	75.439
LB	0.901	53.454	0.934	71.878	0.945	154.819
NewLB	0.982*	43.384	0.984*	60.899	0.981*	137.706

Estimator	$N = 10,000$		$N = 50,000$		$N = 100,000$	
	$1-\hat{\alpha}$	Average Length	$1-\hat{\alpha}$	Average Length	$1-\hat{\alpha}$	Average Length
GT	0.883	104.895	0.888	232.061	0.895	327.412
NewGT	0.818	105.522	0.842	232.631	0.841	328.009
LB	0.944	218.117	0.949*	484.120	0.938	685.143
NewLB	0.987*	197.567	0.991*	446.143	0.990*	641.387

* หมายถึง ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นไม่น้อยกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด หรือ $1-\hat{\alpha} \geq 0.9464$

Ref. code: 25595809030157UGW

ตารางที่ 4.11 แสดงค่าประมาณสัมประสิทธิ์ความเชื่อมั่นและค่าประมาณความกว้างเฉลี่ยของช่วงเชื่อมั่นของตัวประมาณทั้ง 4 ภายใต้การแจกแจงปัวซง $p_i \sim Poi(\lambda)$ เมื่อกำหนดให้ $\lambda = 5$ และขนาดประชากร $N = 500, 1,000, 5,000 ,10,000 ,50,000 ,100,000$

Estimator	$N=500$		$N=1,000$		$N=5,000$	
	$1-\hat{\alpha}$	Average Length	$1-\hat{\alpha}$	Average Length	$1-\hat{\alpha}$	Average Length
GT	0.956*	8.847	0.951*	12.095	0.944	25.891
NewGT	0.927	9.162	0.938	12.072	0.948*	25.812
LB	0.899	18.264	0.920	22.370	0.946*	44.153
NewLB	0.962*	9.384	0.957*	12.957	0.960*	27.729

Estimator	$N=10,000$		$N=50,000$		$N=100,000$	
	$1-\hat{\alpha}$	Average Length	$1-\hat{\alpha}$	Average Length	$1-\hat{\alpha}$	Average Length
GT	0.941	36.241	0.949*	79.843	0.946*	112.520
NewGT	0.931	36.105	0.950*	79.740	0.943	112.385
LB	0.950*	61.332	0.941	135.636	0.951*	192.026
NewLB	0.957*	39.186	0.961*	87.474	0.965*	123.906

* หมายถึง ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นไม่น้อยกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด หรือ $1-\hat{\alpha} \geq 0.9464$

Ref. code: 25595809030157UGW

ตารางที่ 4.12 แสดงค่าประมาณสัมประสิทธิ์ความเชื่อมั่นและค่าประมาณความกว้างเฉลี่ยของช่วงเชื่อมั่นของตัวประมาณทั้ง 4 ภายใต้การแจกแจงปัวซง $p_i \sim Poi(\lambda)$ เมื่อกำหนดให้ $\lambda = 7$ และขนาดประชากร $N = 500, 1,000, 5,000 ,10,000 ,50,000 ,100,000$

Estimator	$N = 500$		$N = 1,000$		$N = 5,000$	
	$1-\hat{\alpha}$	Average Length	$1-\hat{\alpha}$	Average Length	$1-\hat{\alpha}$	Average Length
GT	0.941	3.178	0.955*	4.418	0.953*	9.393
NewGT	0.936	3.383	0.955*	4.369	0.967*	9.238
LB	0.879	12.220	0.879	9.631	0.918	14.766
NewLB	0.906	2.987	0.923	4.123	0.957*	9.163

Estimator	$N = 10,000$		$N = 50,000$		$N = 100,000$	
	$1-\hat{\alpha}$	Average Length	$1-\hat{\alpha}$	Average Length	$1-\hat{\alpha}$	Average Length
GT	0.962*	13.146	0.959*	28.783	0.946*	40.503
NewGT	0.963*	12.938	0.946*	28.601	0.950*	40.296
LB	0.940	20.284	0.951*	43.704	0.951*	61.597
NewLB	0.968*	12.887	0.959*	28.839	0.947*	40.785

* หมายถึง ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นไม่น้อยกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด หรือ $1-\hat{\alpha} \geq 0.9464$

Ref. code: 25595809030157UGW

ตารางที่ 4.13 แสดงค่าประมาณสัมประสิทธิ์ความเชื่อมั่นและค่าประมาณความกว้างเฉลี่ยของช่วงเชื่อมั่นของตัวประมาณทั้ง 4 ภายใต้การแจกแจงปัวซงแบบผสมที่มีพารามิเตอร์ λ_1, λ_2 หรือ $p_i \sim 0.5Poi(\lambda_1) + 0.5Poi(\lambda_2)$ โดยที่ $(\lambda_1, \lambda_2) = (3,5)$ และ $N = 500, 1,000, 5,000 ,10,000 ,50,000 ,100,000$

Estimator	$N = 500$		$N = 1,000$		$N = 5,000$	
	$1-\hat{\alpha}$	Average Length	$1-\hat{\alpha}$	Average Length	$1-\hat{\alpha}$	Average Length
GT	0.892	16.559	0.905	23.353	0.906	53.264
NewGT	0.850	17.813	0.862	24.484	0.826	54.132
LB	0.916	39.855	0.943	53.827	0.951*	112.677
NewLB	0.952*	23.521	0.964*	32.766	0.950*	71.624

Estimator	$N = 10,000$		$N = 50,000$		$N = 100,000$	
	$1-\hat{\alpha}$	Average Length	$1-\hat{\alpha}$	Average Length	$1-\hat{\alpha}$	Average Length
GT	0.888	75.695	0.858	170.305	0.773	240.953
NewGT	0.823	76.685	0.668	172.149	0.501	243.443
LB	0.955*	157.728	0.955*	351.459	0.944	495.197
NewLB	0.932	101.559	0.902	244.621	0.839	369.129

* หมายถึง ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นไม่น้อยกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด หรือ $1-\hat{\alpha} \geq 0.9464$

Ref. code: 25595809030157UGW

ตารางที่ 4.14 แสดงค่าประมาณสัมประสิทธิ์ความเชื่อมั่นและค่าประมาณความกว้างเฉลี่ยของช่วงเชื่อมั่นของตัวประมาณทั้ง 4 ภายใต้การแจกแจงปัวซงแบบผสมที่มีพารามิเตอร์ λ_1, λ_2 หรือ $p_i \sim 0.5Poi(\lambda_1) + 0.5Poi(\lambda_2)$ โดยที่ $(\lambda_1, \lambda_2) = (7,5)$ และ $N = 500, 1,000, 5,000 ,10,000 ,50,000 ,100,000$

Estimator	$N=500$		$N=1,000$		$N=5,000$	
	$1-\hat{\alpha}$	Average Length	$1-\hat{\alpha}$	Average Length	$1-\hat{\alpha}$	Average Length
GT	0.874	13.951	0.879	20.421	0.659	35.165
NewGT	0.772	12.827	0.745	27.856	0.370	19.986
LB	0908	38.749	0.943	52.022	0.962*	108.593
NewLB	0.971*	21.976	0.969*	30.093	0.974*	71.458

Estimator	$N=10,000$		$N=50,000$		$N=100,000$	
	$1-\hat{\alpha}$	Average Length	$1-\hat{\alpha}$	Average Length	$1-\hat{\alpha}$	Average Length
GT	0.364	27.561	0.000	0.000	0.000	0.000
NewGT	0.137	10.393	0.001	0.161	0.000	0.000
LB	0.963*	151.992	0.923	321.556	0.914	449.425
NewLB	0.997*	116.556	1.000*	422.758	1.000*	802.457

* หมายถึง ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นไม่น้อยกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด หรือ $1-\hat{\alpha} \geq 0.9464$

Ref. code: 25595809030157UGW

ตารางที่ 4.15 แสดงค่าประมาณสัมประสิทธิ์ความเชื่อมั่นและค่าประมาณความกว้างเฉลี่ยของช่วงเชื่อมั่นของตัวประมาณทั้ง 4 ภายใต้การแจกแจงทวินามเชิงลบที่มีพารามิเตอร์ π, κ , $p_i \sim NB(\kappa, \pi)$ โดยที่ $\kappa = 4, \pi = 0.5$ และขนาดประชากร $N = 500, 1,000, 5,000, 10,000, 50,000, 100,000$

Estimator	$N=500$		$N=1,000$		$N=5,000$	
	$1-\hat{\alpha}$	Average Length	$1-\hat{\alpha}$	Average Length	$1-\hat{\alpha}$	Average Length
GT	0.699	16.969	0.703	24.621	0.754	60.961
NewGT	0.630	15.879	0.608	22.006	0.605	49.517
LB	0.910	70.441	0.917	92.698	0.925	194.153
NewLB	0.867	35.591	0.807	46.685	0.491	76.219

Estimator	$N=10,000$		$N=50,000$		$N=100,000$	
	$1-\hat{\alpha}$	Average Length	$1-\hat{\alpha}$	Average Length	$1-\hat{\alpha}$	Average Length
GT	0.773	88.966	0.802	207.799	0.816	299.474
NewGT	0.603	69.905	0.592	153.783	0.597	219.633
LB	0.922	270.977	0.872	569.393	0.801	739.767
NewLB	0.303	79.136	0.029	31.175	0.003	6.948

* หมายถึง ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นไม่น้อยกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด หรือ $1-\hat{\alpha} \geq 0.9464$

Ref. code: 25595809030157UGW

ตารางที่ 4.16 แสดงค่าประมาณสัมประสิทธิ์ความเชื่อมั่นและค่าประมาณความกว้างเฉลี่ยของช่วงเชื่อมั่นของตัวประมาณทั้ง 4 ภายใต้การแจกแจงทวินามเชิงลบที่มีพารามิเตอร์ π, κ , $p_i \sim NB(\kappa, \pi)$ โดยที่ $\kappa = 4, \pi = 0.8$ และขนาดประชากร $N = 500, 1,000, 5,000 ,10,000 ,50,000 ,100,000$

Estimator	$N = 500$		$N = 1,000$		$N = 5,000$	
	$1-\hat{\alpha}$	Average Length	$1-\hat{\alpha}$	Average Length	$1-\hat{\alpha}$	Average Length
GT	0.628	72.169	0.568	90.956	0.548	197.636
NewGT	0.579	66.861	0.503	81.427	0.369	133.978
LB	0.905	359.474	0.919	482.329	0.940	1035.938
NewLB	0.985*	577.631	0.976*	909.247	0.958*	2464.099

Estimator	$N = 10,000$		$N = 50,000$		$N = 100,000$	
	$1-\hat{\alpha}$	Average Length	$1-\hat{\alpha}$	Average Length	$1-\hat{\alpha}$	Average Length
GT	0.549	279.826	0.574	657.025	0.581	940.808
NewGT	0.307	157.772	0.316	362.382	0.286	463.645
LB	0.914	1416.918	0.843	2909.143	0.730	3567.072
NewLB	0.911	3803.883	0.628	8531.709	0.408	9758.185

* หมายถึง ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นไม่น้อยกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด หรือ $1-\hat{\alpha} \geq 0.9464$

Ref. code: 25595809030157UGW

ตารางที่ 4.17 แสดงค่าประมาณสัมประสิทธิ์ความเชื่อมั่นและค่าประมาณความกว้างเฉลี่ยของช่วงเชื่อมั่นของตัวประมาณทั้ง 4 ภายใต้การแจกแจงทวินามเชิงลบที่มีพารามิเตอร์ π, κ , $p_i \sim NB(\kappa, \pi)$ โดยที่ $\kappa = 9, \pi = 0.5$ และขนาดประชากร $N = 500, 1,000, 5,000 ,10,000 ,50,000 ,100,000$

Estimator	$N = 500$		$N = 1,000$		$N = 5,000$	
	$1-\hat{\alpha}$	Average Length	$1-\hat{\alpha}$	Average Length	$1-\hat{\alpha}$	Average Length
GT	0.905	3.685	0.908	5.188	0.938	12.207
NewGT	0.888	4.203	0.913	5.732	0.935	12.521
LB	0.855	26.968	0.877	20.041	0.922	25.563
NewLB	0.904	4.721	0.937	5.899	0.918	11.668

Estimator	$N = 10,000$		$N = 50,000$		$N = 100,000$	
	$1-\hat{\alpha}$	Average Length	$1-\hat{\alpha}$	Average Length	$1-\hat{\alpha}$	Average Length
GT	0.947*	17.524	0.931	38.980	0.933	55.386
NewGT	0.952*	17.973	0.924	38.899	0.923	54.910
LB	0.940	33.757	0.941	71.848	0.943	101.075
NewLB	0.885	15.810	0.628	25.640	0.414	24.599

* หมายถึง ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นไม่น้อยกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด หรือ $1-\hat{\alpha} \geq 0.9464$

Ref. code: 25595809030157UGW

ตารางที่ 4.18 แสดงค่าประมาณสัมประสิทธิ์ความเชื่อมั่นและค่าประมาณความกว้างเฉลี่ยของช่วงเชื่อมั่นของตัวประมาณทั้ง 4 ภายใต้การแจกแจงทวินามเชิงลบที่มีพารามิเตอร์ π, κ , $p_i \sim NB(\kappa,\pi)$ โดยที่ $\kappa = 9, \pi = 0.8$ และขนาดประชากร $N = 500, 1{,}000, 5{,}000, 10{,}000, 50{,}000, 100{,}000$

Estimator	$N=500$		$N=1{,}000$		$N=5{,}000$	
	$1-\hat{\alpha}$	Average Length	$1-\hat{\alpha}$	Average Length	$1-\hat{\alpha}$	Average Length
GT	0.782	31.891	0.748	43.142	0.744	97.346
NewGT	0.725	30.182	0.687	40.341	0.587	77.359
LB	0.922	114.258	0.926	150.195	0.940	323.951
NewLB	0.974*	115.275	0.968*	164.927	0.949*	410.521

Estimator	$N=10{,}000$		$N=50{,}000$		$N=100{,}000$	
	$1-\hat{\alpha}$	Average Length	$1-\hat{\alpha}$	Average Length	$1-\hat{\alpha}$	Average Length
GT	0.777	144.112	0.788	328.485	0.817	482.061
NewGT	0.637	118.833	0.631	263.420	0.628	370.802
LB	0.939	456.543	0.944	1019.489	0.935	1428.667
NewLB	0.877	551.730	0.505	928.919	0.226	701.312

* หมายถึง ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นไม่น้อยกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด หรือ $1-\hat{\alpha} \geq 0.9464$

Ref. code: 25595809030157UGW

4.3 การประยุกต์ใช้กับข้อมูลจริง

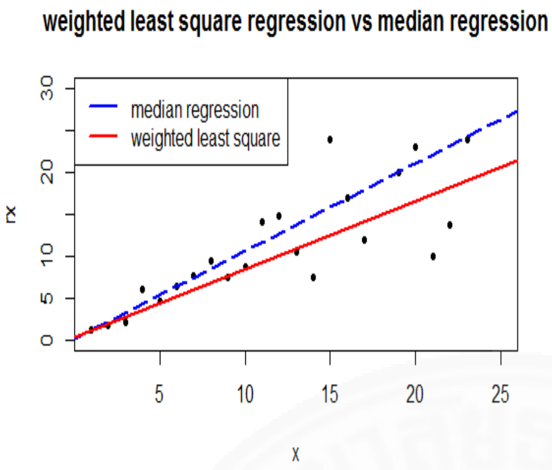
เพื่อแสดงให้เห็นถึงประสิทธิภาพของตัวประมาณขนาดประชากรและช่วงเชื่อมั่นของตัวประมาณขนาดประชากรที่ปรับปรุงโดยใช้การถดถอยควอนไทล์ที่ผู้วิจัยได้นำเสนอ จึงได้นำตัวประมาณขนาดประชากรที่ได้พัฒนาขึ้นใหม่นั้นมาประยุกต์ใช้กับข้อมูลจริง ซึ่งในที่นี้ เราสนใจการประมาณจำนวนสปีชีส์ของผีเสื้อมลายูในเกาะเตียวมัน ประเทศมาเลเซีย แสดงข้อมูลได้ดังตารางที่ 4.19 ศึกษารายละเอียดเพิ่มเติมของข้อมูลได้ใน Fisher, Corbet and Williams(1943) และกราฟแสดงความสัมพันธ์ระหว่าง log-ratios $\log(r_x)$ กับ x ของการถดถอยเชิงเส้นแบบถ่วงน้ำหนัก และการถดถอยควอนไทล์ ที่ $\tau = 0.50$ แสดงได้ดังภาพที่ 4.1

ตารางที่ 4.19 จำนวนครั้งและจำนวนความถี่ที่พบจำนวนสปีชีส์ของผีเสื้อมลายู ในเกาะเตียวมัน ประเทศมาเลเซีย ซึ่งจำนวนของจำนวนสปีชีส์ของผีเสื้อมลายู รวมทั้งสิ้น 620 สปีชีส์

x	1	2	3	4	5	6	7	8	9	10	11	12	
f_x	118	74	44	24	29	22	20	19	20	15	12	14	
x	13	14	15	16	17	18	19	20	21	22	23	24	25>
f_x	16	12	6	9	9	6	10	10	11	5	3	3	119

จากตารางที่ 4.19 พบว่า มีสปีชีส์ของผีเสื้อมลายูจำนวน 118 สปีชีส์ ที่ถูกพบ 1 ครั้ง สปีชีส์ของผีเสื้อมลายูจำนวน 74 สปีชีส์ ที่ถูกพบ 2 ครั้ง และสปีชีส์ของผีเสื้อมลายูจำนวน 44 สปีชีส์ ที่ถูกพบ 3 ครั้ง เป็นต้น จะเห็นได้ว่า จำนวนสปีชีส์ของผีเสื้อมลายูที่ถูกพบรวมทั้งสิ้น 620 สปีชีส์ นั่นคือ $n = 118 + 74 + 44 + 24 + 29 + ... + 3 + 119 = 620$ จาก $N = n + f_0$ เมื่อ f_0 คือ จำนวนสปีชีส์ของผีเสื้อมลายูที่ไม่สุ่มพบ ทำให้ต้องประมาณจำนวนสปีชีส์ของผีเสื้อมลายูที่ไม่สุ่มพบ เพื่อประมาณจำนวนสปีชีส์ของผีเสื้อมลายูในเกาะเตียวมัน ประเทศมาเลเซีย

Ref. code: 25595809030157UGW



ภาพที่ 4.1(ก) แสดงกราฟแสดงความสัมพันธ์ระหว่าง r_x กับ x ของการถดถอยเชิงเส้นแบบถ่วงน้ำหนักและการถดถอยควอนไทล์ที่ $\tau = 0.50$

ภาพที่ 4.1(ข) แสดงกราฟแสดงความสัมพันธ์ระหว่าง $\log(r_x)$ กับ x ของการถดถอยเชิงเส้นแบบถ่วงน้ำหนักและการถดถอยควอนไทล์ที่ $\tau = 0.50$

จากภาพที่ 4.1(ก) และ (ข) เห็นได้ว่า เมื่อเปรียบเทียบเส้นถดถอยของการถดถอยเชิงเส้นแบบถ่วงน้ำหนัก และการถดถอยควอนไทล์ที่ระดับ 0.50 ($\tau = 0.50$) พบว่า เส้นตรงมีลักษณะที่มีความชันเป็นบวก หรือเป็นลักษณะของ heterogeneous poisson model และพบว่าเส้นถดถอยของการถดถอยควอนไทล์ ที่ $\tau = 0.50$ หรือการถดถอยเข้าสู่ค่ามัธยฐานนั้นมีค่าใกล้เคียงกับชุดข้อมูลมากกว่าเส้นถดถอยของการถดถอยเชิงเส้นแบบถ่วงน้ำหนัก ดังนั้นเมื่อประมาณค่าพารามิเตอร์จากการถดถอยที่เข้าสู่ค่ามัธยฐานควรจะให้ตัวประมาณค่าพารามิเตอร์ที่ดี และทำให้ประมาณค่าขนาดประชากรดีขึ้น เพื่อแสดงประสิทธิภาพของตัวประมาณที่ปรับปรุงโดยใช้การถดถอยควอนไทล์ แสดงผลที่ได้ดังตารางที่ 4.20

Ref. code: 25595809030157UGW

ตารางที่ 4.20 แสดงค่าประมาณขนาดประชากร, ค่าคลาดเคลื่อนมาตรฐานของตัวประมาณขนาดประชากร และช่วงเชื่อมั่น 95% ของขนาดประชากรสปีชีส์ของผีเสื้อมลายู

Estimator	$\hat{N}$	$SE(\hat{N})$	$95\%CI$
GT	952.09	26.45	900.25-1003.93
NewGT	636.80	4.38	638.22-645.38
LB	753.80	63.49	629.36-878.24
NewLB	655.66	15.01	626.23-685.08

จากตาราง 4.20 พบว่า $\hat{N}_{GT}$ ให้ค่าประมาณขนาดประชากรสูงที่สุด $\hat{N}_{NewGT}$ และ $\hat{N}_{NewLB}$ ให้ค่าประมาณขนาดประชากรแตกต่างกันเพียงเล็กน้อย โดยที่ $\hat{N}_{NewGT}$ ให้ค่าประมาณขนาดประชากรต่ำที่สุดและให้ค่าความคลาดเคลื่อนมาตรฐานต่ำที่สุด อีกทั้งยังให้ช่วงเชื่อมั่นของขนาดประชากรที่แคบที่สุด ดังนั้นหากเปรียบเทียบประสิทธิภาพระหว่าง $\hat{N}_{GT}$ และ สรุปได้ว่า $\hat{N}_{NewGT}$ เป็นตัวเลือกที่เหมาะสมกว่า $\hat{N}_{GT}$

เมื่อเปรียบเทียบประสิทธิภาพระหว่าง $\hat{N}_{LB}$ และ $\hat{N}_{NewLB}$ พบว่า $\hat{N}_{NewLB}$ ให้ค่าประมาณขนาดประชากรต่ำกว่า $\hat{N}_{LB}$ และพบว่า $\hat{N}_{NewLB}$ ให้ค่าความคลาดเคลื่อนมาตรฐานต่ำกว่า อีกทั้งยังให้ช่วงเชื่อมั่นของขนาดประชากรที่แคบกว่าของ $\hat{N}_{LB}$ จึงถือได้ว่า $\hat{N}_{NewLB}$ เป็นตัวเลือกที่เหมาะสมกว่า $\hat{N}_{LB}$

บทที่ 5

สรุปผลการศึกษาและข้อเสนอแนะ

การศึกษาครั้งนี้มีวัตถุประสงค์เพื่อศึกษาเปรียบเทียบประสิทธิภาพของตัวประมาณขนาดประชากร และประสิทธิภาพของช่วงเชื่อมั่นของตัวประมาณขนาดประชากรทั้ง 4 คือ $\hat{N}_{GT}, \hat{N}_{NewGT}, \hat{N}_{LB}$ และ $\hat{N}_{NewLB}$ โดยได้สรุปผลการศึกษาเป็น 7 หัวข้อ คือ

5.1 ผลการศึกษาจากค่าความเอนเอียงสัมพัทธ์ ($Rbias$)

5.2 ผลการศึกษาจากค่าความแปรปรวนสัมพัทธ์ ($RVar$)

5.3 ผลการศึกษาจากค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ ($RMSE$)

5.4 ผลการศึกษาจากค่าประมาณสัมประสิทธิ์ความเชื่อมั่นของขนาดประชากร $(1-\hat{\alpha})$

5.5 ผลการศึกษาจากค่าประมาณความกว้างเฉลี่ยของช่วงเชื่อมั่น (Average Length)

5.6 ผลการศึกษาเปรียบเทียบประสิทธิภาพของตัวประมาณขนาดประชากร

5.7 ผลการศึกษาเปรียบเทียบประสิทธิภาพของช่วงเชื่อมั่นของขนาดประชากร

ผลการเปรียบเทียบประสิทธิภาพของตัวประมาณขนาดประชากรทั้ง 4 ตัวคือ $\hat{N}_{GT}, \hat{N}_{NewGT}, \hat{N}_{LB}$ และ $\hat{N}_{NewLB}$ นำเสนอด้วยตารางที่ 5.1-5.3 ที่แสดงค่าความเอนเอียงสัมพัทธ์ ($RBias$) ค่าความแปรปรวนสัมพัทธ์ ($RVar$) และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ ($RMSE$) สำหรับการจำลองภายใต้ตัวแบบการแจกแจงปัวซงที่มีพารามิเตอร์ λ การจำลองภายใต้ตัวแบบการแจกแจงปัวซงแบบผสมที่มีพารามิเตอร์ λ_1, λ_2 และการจำลองภายใต้ตัวแบบการแจกแจงทวินามเชิงลบที่มีพารามิเตอร์ κ, π ตามลำดับ และผลการเปรียบเทียบประสิทธิภาพของช่วงเชื่อมั่นของขนาดประชากรที่สร้างจากตัวประมาณขนาดประชากรทั้ง 4 ตัวคือ $\hat{N}_{GT}, \hat{N}_{NewGT}, \hat{N}_{LB}$ และ $\hat{N}_{NewLB}$ นำเสนอด้วยตารางที่ 5.4-5.5 ที่แสดงค่าประมาณสัมประสิทธิ์ความเชื่อมั่นของขนาดประชากร $(1-\hat{\alpha})$ และค่าประมาณความกว้างเฉลี่ยของช่วงเชื่อมั่น (Average Length) สำหรับการจำลองภายใต้ตัวแบบการแจกแจงปัวซงที่มีพารามิเตอร์ λ การจำลองภายใต้ตัวแบบการแจกแจงปัวซงแบบผสมที่มีพารามิเตอร์ λ_1, λ_2 และการจำลองภายใต้ตัวแบบการแจกแจงทวินามเชิงลบที่มีพารามิเตอร์ κ, π ตามลำดับ

Ref. code: 25595809030157UGW

5.1 ผลการศึกษาจากค่าความเอนเอียงสัมพัทธ์ $(RBias)$

ตารางที่ 5.1 แสดงผลสรุปตัวประมาณขนาดประชากรที่มีประสิทธิสูงสุดเมื่อพิจารณาค่าความเอนเอียงสัมพัทธ์ $(RBias)$

N / $Model$	500	1,000	5,000	10,000	50,000	100,000
$p_i \sim Poi(\lambda)$						
$\lambda = 3$	$\hat{N}_{NewLB}$	$\hat{N}_{NewLB}$	$\hat{N}_{NewLB}$	$\hat{N}_{NewLB}$	$\hat{N}_{NewLB}$	$\hat{N}_{NewLB}$
$\lambda = 5$	$\hat{N}_{NewLB}$	$\hat{N}_{NewLB}$	$\hat{N}_{NewLB}$	$\hat{N}_{NewLB}$	$\hat{N}_{NewLB}$	$\hat{N}_{NewLB}$
$\lambda = 7$	$\hat{N}_{NewLB}$	$\hat{N}_{NewLB}$	$\hat{N}_{NewLB}$	$\hat{N}_{NewLB}$	$\hat{N}_{NewLB}$	$\hat{N}_{NewLB}$
$p_i \sim 0.5Poi(\lambda_1)+0.5Poi(\lambda_2)$						
$(\lambda_1,\lambda_2)=(3,5)$	$\hat{N}_{GT}$	$\hat{N}_{NewLB}$	$\hat{N}_{GT}$	$\hat{N}_{GT}$	$\hat{N}_{LB}$	$\hat{N}_{LB}$
$(\lambda_1,\lambda_2)=(7,5)$	$\hat{N}_{GT}$	$\hat{N}_{NewLB}$	$\hat{N}_{NewLB}$	$\hat{N}_{NewLB}$	$\hat{N}_{NewLB}$	$\hat{N}_{NewLB}$
$p_i \sim NB(\kappa,\pi)$						
$\kappa=4,\pi=0.5$	$\hat{N}_{NewGT}$	$\hat{N}_{NewGT}$	$\hat{N}_{NewGT}$	$\hat{N}_{NewGT}$	$\hat{N}_{NewGT}$	$\hat{N}_{NewGT}$
$\kappa=4,\pi=0.8$	$\hat{N}_{GT}$	$\hat{N}_{LB}$	$\hat{N}_{NewGT}$	$\hat{N}_{NewGT}$	$\hat{N}_{NewGT}$	$\hat{N}_{NewGT}$
$\kappa=9,\pi=0.5$	$\hat{N}_{GT}$	$\hat{N}_{NewLB}$	$\hat{N}_{NewGT}$	$\hat{N}_{NewGT}$	$\hat{N}_{NewGT}$	$\hat{N}_{NewGT}$
$\kappa=9,\pi=0.8$	$\hat{N}_{GT}$	$\hat{N}_{NewLB}$	$\hat{N}_{LB}$	$\hat{N}_{NewGT}$	$\hat{N}_{NewGT}$	$\hat{N}_{NewGT}$

จากตารางที่ 5.1 เมื่อพิจารณาจากค่าความเอนเอียงสัมพัทธ์ $(RBias)$ ภายใต้การแจกแจงปัวซงที่มีพารามิเตอร์ $\lambda = 3,5,7$ ภายใต้การแจกแจงปัวซงแบบผสมที่มีพารามิเตอร์ λ_1,λ_2 ที่มีพารามิเตอร์ $(\lambda_1,\lambda_2)=(3,5)$ และ $(\lambda_1,\lambda_2)=(7,5)$ และ ภายใต้การแจกแจงทวินามเชิงลบ ที่มีพารามิเตอร์ κ,π, $p_i \sim NB(\kappa,\pi)$ โดยที่ $\kappa=(4,9)$, $\pi=(0.5,0.8)$ สำหรับขนาดประชากรขนาดเล็ก $N=500,1{,}000$ ประชากรขนาดกลาง $N=5{,}000$, $10{,}000$ และประชากรขนาดใหญ่ $N=50{,}000$, $100{,}000$ สามารถสรุปผลการเปรียบเทียบประสิทธิภาพของตัวประมาณขนาดประชากรทั้ง 4 ตัวได้ดังนี้

ภายใต้ตัวแบบการแจกแจงปัวซงที่มีพารามิเตอร์ $\lambda = 3,5,7$ นั้น $\hat{N}_{NewLB}$ ให้ค่า $RBias$ ต่ำสุดทุกขนาดประชากรที่ศึกษาคือ 500, 1,000, 5,000, 10,000, 50,000 และ 100,000

จากผลดังกล่าวข้างต้นสรุปได้ว่า $\hat{N}_{NewLB}$ มีประสิทธิภาพสูงสำหรับการแจกแจงปัวซงที่มีพารามิเตอร์ $\lambda = 3,5,7$ ในทุกขนาดประชากรที่ศึกษา

Ref. code: 25595809030157UGW

ภายใต้ตัวแบบการแจกแจงปัวซงแบบผสมที่มีพารามิเตอร์ λ_1, λ_2 ที่มีพารามิเตอร์ $(\lambda_1, \lambda_2) = (3,5)$ และ $(\lambda_1, \lambda_2) = (7,5)$ สำหรับประชากร N ทั้งหมดที่ศึกษา เมื่อพารามิเตอร์ $(\lambda_1, \lambda_2) = (3,5)$ พบว่า $\hat{N}_{GT}$ ให้ค่า $RBias$ ต่ำสุดเมื่อขนาดประชากรเป็น 500, 5,000 และ 10,000 ส่วน $\hat{N}_{LB}$ ให้ค่า $RBias$ ต่ำสุดเมื่อขนาดประชากรเป็น 50,000 และ 100,000 และ $\hat{N}_{NewLB}$ ให้ค่า $RBias$ ต่ำสุดเมื่อขนาดประชากรเป็น 1,000

เมื่อพารามิเตอร์ $(\lambda_1, \lambda_2) = (7,5)$ นั้น $\hat{N}_{NewLB}$ ให้ค่า $RBias$ ต่ำสุดทุกขนาดประชากรที่ศึกษา ยกเว้นที่ขนาดประชากรเป็น 500 พบว่า $\hat{N}_{GT}$ ให้ค่า $RBias$ ต่ำสุด

จากผลดังกล่าวข้างต้นสรุปได้ว่า $\hat{N}_{GT}$ มีประสิทธิภาพสูงสำหรับการแจกแจงปัวซงแบบผสมที่มีพารามิเตอร์ λ_1, λ_2 ที่มีพารามิเตอร์ $(\lambda_1, \lambda_2) = (3,5)$ และ $(\lambda_1, \lambda_2) = (7,5)$ เมื่อประชากรมีขนาดเล็ก และ $\hat{N}_{NewLB}$ มีประสิทธิภาพสูงสำหรับการแจกแจงปัวซงแบบผสมที่มีพารามิเตอร์ λ_1, λ_2 ที่มีพารามิเตอร์ $(\lambda_1, \lambda_2) = (7,5)$ เมื่อประชากรขนาดกลางและขนาดใหญ่

ภายใต้ตัวแบบการแจกแจงทวินามเชิงลบ ที่มีพารามิเตอร์ π และ κ โดยที่ $\kappa = (4,9), \pi = (0.5, 0.8)$ สำหรับประชากร N ทั้งหมดที่ศึกษา เมื่อ $\kappa = 4, \pi = 0.5$ พบว่า $\hat{N}_{NewGT}$ ให้ค่า $RBias$ ต่ำสุดในทุกขนาดประชากรที่ศึกษา

เมื่อ $\kappa = 4, \pi = 0.8$ พบว่า $\hat{N}_{NewGT}$ ให้ค่า $RBias$ ต่ำสุดเมื่อขนาดประชากรเป็น 5,000, 10,000, 50,000 และ 100,000 และ $\hat{N}_{GT}$ ให้ค่า $RBias$ ต่ำสุดเมื่อขนาดประชากรเป็น 500 และ $\hat{N}_{LB}$ ให้ค่า $RBias$ ต่ำสุดเมื่อขนาดประชากรเป็น 1,000

และ $\kappa = 8, \pi = 0.5$ พบว่า $\hat{N}_{NewGT}$ ให้ค่า $RBias$ ต่ำสุดเมื่อขนาดประชากรเป็น 5,000, 10,000, 50,000 และ 100,000 และ $\hat{N}_{GT}$ ให้ค่า $RBias$ ต่ำสุดเมื่อขนาดประชากรเป็น 500 และ $\hat{N}_{NewLB}$ ให้ค่า $RBias$ ต่ำสุดเมื่อขนาดประชากรเป็น 1,000

เมื่อ $\kappa = 8, \pi = 0.8$ พบว่า $\hat{N}_{NewGT}$ ให้ค่า $RBias$ ต่ำสุดเมื่อขนาดประชากรเป็น 10,000, 50,000 และ 100,000 ส่วน $\hat{N}_{GT}$ ให้ค่า $RBias$ ต่ำสุดเมื่อขนาดประชากรเป็น 500 , $\hat{N}_{NewLB}$ ให้ค่า $RBias$ ต่ำสุดเมื่อขนาดประชากรเป็น 1,000 และ $\hat{N}_{LB}$ ให้ค่า $RBias$ ต่ำสุดเมื่อขนาดประชากรเป็น 5,000

จากผลดังกล่าวข้างต้นสรุปได้ว่า $\hat{N}_{NewGT}$ มีประสิทธิภาพสูงสำหรับการแจกแจงทวินามเชิงลบ ที่มีพารามิเตอร์ π และ κ โดยที่ $\kappa = 4, \pi = 0.5$ ในทุกขนาดประชากรที่ศึกษา และเมื่อ $\kappa = 8, \pi = (0.5, 0.8)$ $\hat{N}_{NewGT}$ มีประสิทธิภาพสูงเมื่อประชากรมีขนาดกลางและขนาดใหญ่

Ref. code: 25595809030157UGW

5.2 ผลการศึกษาจากค่าความแปรปรวนสัมพัทธ์ ($RVar$)

ตารางที่ 5.2 แสดงผลสรุปตัวประมาณขนาดประชากรที่มีประสิทธิสูงสุดเมื่อพิจารณาค่าความแปรปรวนสัมพัทธ์ ($RVar$)

Model \ N	500	1,000	5,000	10,000	50,000	100,000
$p_i \sim Poi(\lambda)$						
$\lambda = 3$	$\hat{N}_{GT}, \hat{N}_{NewLB}$	$\hat{N}_{GT}, \hat{N}_{NewLB}$	$\hat{N}_{GT}, \hat{N}_{NewLB}$	$\hat{N}_{GT}, \hat{N}_{NewLB}$	$\hat{N}_{GT}, \hat{N}_{NewLB}$	$\hat{N}_{GT}, \hat{N}_{NewLB}$
$\lambda = 5$	$\hat{N}_{GT}, \hat{N}_{NewLB}$	$\hat{N}_{GT}, \hat{N}_{NewLB}$	$\hat{N}_{GT}, \hat{N}_{NewLB}$	$\hat{N}_{GT}, \hat{N}_{NewLB}$	$\hat{N}_{GT}, \hat{N}_{NewLB}$	$\hat{N}_{GT}, \hat{N}_{NewLB}$
$\lambda = 7$	$\hat{N}_{GT}, \hat{N}_{NewLB}$	$\hat{N}_{GT}, \hat{N}_{NewLB}$	$\hat{N}_{GT}, \hat{N}_{NewLB}$	$\hat{N}_{GT}, \hat{N}_{NewLB}$	$\hat{N}_{GT}, \hat{N}_{NewLB}$	$\hat{N}_{GT}, \hat{N}_{NewLB}$
$p_i \sim 0.5Poi(\lambda_1) + 0.5Poi(\lambda_2)$						
$(\lambda_1, \lambda_2) = (3,5)$	$\hat{N}_{GT}, \hat{N}_{NewLB}$	$\hat{N}_{GT}$	$\hat{N}_{GT}$	$\hat{N}_{NewLB}$	$\hat{N}_{GT}$	$\hat{N}_{NewLB}$
$(\lambda_1, \lambda_2) = (7,5)$	$\hat{N}_{GT}, \hat{N}_{NewLB}$	$\hat{N}_{NewLB}$	$\hat{N}_{NewLB}$	$\hat{N}_{NewLB}$	$\hat{N}_{NewLB}$	$\hat{N}_{GT}$
$p_i \sim NB(\kappa, \pi)$						
$\kappa = 4, \pi = 0.5$	$\hat{N}_{NewLB}$	$\hat{N}_{NewLB}$	$\hat{N}_{NewLB}$	$\hat{N}_{NewLB}$	$\hat{N}_{NewLB}$	$\hat{N}_{NewLB}$
$\kappa = 4, \pi = 0.8$	$\hat{N}_{NewLB}$	$\hat{N}_{NewLB}$	$\hat{N}_{NewLB}$	$\hat{N}_{NewLB}$	$\hat{N}_{NewLB}$	$\hat{N}_{NewLB}$
$\kappa = 9, \pi = 0.5$	$\hat{N}_{NewLB}$	$\hat{N}_{NewLB}$	$\hat{N}_{NewLB}$	$\hat{N}_{NewLB}$	$\hat{N}_{NewLB}$	$\hat{N}_{NewLB}$
$\kappa = 9, \pi = 0.8$	$\hat{N}_{NewLB}$	$\hat{N}_{NewLB}$	$\hat{N}_{NewLB}$	$\hat{N}_{NewLB}$	$\hat{N}_{NewLB}$	$\hat{N}_{NewLB}$

จากตารางที่ 5.2 เมื่อพิจารณาจากค่าความแปรปรวนสัมพัทธ์ ($RVar$) ภายใต้การแจกแจงปัวซงที่มีพารามิเตอร์ $\lambda = 3,5,7$ ภายใต้การแจกแจงปัวซงแบบผสมที่มีพารามิเตอร์ λ_1, λ_2 ที่มีพารามิเตอร์ $(\lambda_1, \lambda_2) = (3,5)$ และ $(\lambda_1, \lambda_2) = (7,5)$ และ ภายใต้การแจกแจงทวินามเชิงลบ ที่มีพารามิเตอร์ κ, π, $p_i \sim NB(\kappa, \pi)$ โดยที่ $\kappa = (4,9), \pi = (0.5, 0.8)$ สำหรับขนาดประชากรขนาดเล็ก $N = 500, 1{,}000$ ประชากรขนาดกลาง $N = 5{,}000,\ 10{,}000$ และประชากรขนาดใหญ่ $N = 50{,}000,\ 100{,}000$ สามารถสรุปผลการเปรียบเทียบประสิทธิภาพของตัวประมาณขนาดประชากรทั้ง 4 ตัวได้ดังนี้

เมื่ออยู่ภายใต้ตัวแบบการแจกแจงปัวซงที่มีพารามิเตอร์ $\lambda = 3,5,7$ สำหรับขนาดประชากร $N = 500, 1{,}000, 5{,}000,\ 10{,}000,\ 50{,}000, 100{,}000$ พบว่า $\hat{N}_{GT}$ และ $\hat{N}_{NewLB}$ มีค่า $RVar$ ต่ำที่สุดในทุกขนาดประชากรที่ศึกษา

จากผลดังกล่าวข้างต้นสรุปได้ว่า $\hat{N}_{GT}$ และ $\hat{N}_{NewLB}$ มีประสิทธิภาพสูงสำหรับการแจกแจงปัวซงที่มีพารามิเตอร์ $\lambda = 3,5,7$ ในทุกขนาดประชากรที่ศึกษา

Ref. code: 25595809030157UGW

เมื่ออยู่ภายใต้ตัวแบบการแจกแจงปัวซงแบบผสมที่มีพารามิเตอร์ λ_1,λ_2 ที่มีพารามิเตอร์ $(\lambda_1,\lambda_2)=(3,5)$ และ $(\lambda_1,\lambda_2)=(7,5)$ สำหรับ $N=500, 1,000, 5,000, 10,000, 50,000, 100,000$ พบว่า $\hat{N}_{GT}$ และ $\hat{N}_{NewLB}$ มีประสิทธิภาพพอๆ กัน แต่เมื่อพารามิเตอร์ λ มีค่าสูงขึ้น จะเห็นได้ว่า $\hat{N}_{NewLB}$ จะมีค่า $RVar$ ต่ำที่สุดเกือบทุกขนาดประชากรที่ศึกษา

เมื่ออยู่ภายใต้ตัวแบบการแจกแจงทวินามเชิงลบ ที่มีพารามิเตอร์ π และ κ โดยที่ $\kappa=(4,9), \pi=(0.5,0.8)$ สำหรับขนาดประชากร $N=500,1,000,5,000, 10,000,50,000,100,000$ พบว่า $\hat{N}_{NewLB}$ ให้ค่า $RVar$ ต่ำสุดและค่ายิ่งเข้าใกล้ค่า 0 มากขึ้นเมื่อประชากรเพิ่มขึ้น ในขนาดประชากรที่ศึกษา

จากผลดังกล่าวข้างต้นสรุปได้ว่า $\hat{N}_{NewLB}$ มีประสิทธิภาพสูงสำหรับการแจกแจงทวินามเชิงลบที่มีพารามิเตอร์ π และ κ โดยที่ $\kappa=(4,9), \pi=(0.5,0.8)$ ในทุกขนาดประชากรที่ศึกษา

5.3 ผลการศึกษาจากค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ ($RMSE$)

ตารางที่ 5.3 แสดงผลสรุปตัวประมาณขนาดประชากรที่มีประสิทธิสูงสุดเมื่อพิจารณาค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ ($RMSE$)

Model \ N	500	1,000	5,000	10,000	50,000	100,000
$p_i \sim Poi(\lambda)$						
$\lambda=3$	$\hat{N}_{GT}, \hat{N}_{NewLB}$	$\hat{N}_{GT}, \hat{N}_{NewLB}$	$\hat{N}_{NewLB}$	$\hat{N}_{GT}, \hat{N}_{NewLB}$	$\hat{N}_{GT}$	$\hat{N}_{GT}$
$\lambda=5$	$\hat{N}_{NewLB}$	$\hat{N}_{NewLB}$	$\hat{N}_{NewLB}$	$\hat{N}_{NewLB}$	$\hat{N}_{GT}$	$\hat{N}_{NewLB}$
$\lambda=7$	$\hat{N}_{GT}, \hat{N}_{NewLB}$	$\hat{N}_{NewLB}$	$\hat{N}_{NewGT}$	$\hat{N}_{NewLB}$	$\hat{N}_{NewLB}$	$\hat{N}_{NewLB}$
$p_i \sim 0.5Poi(\lambda_1)+0.5Poi(\lambda_2)$						
$(\lambda_1,\lambda_2)=(3,5)$	$\hat{N}_{GT}, \hat{N}_{NewLB}$	$\hat{N}_{GT}$	$\hat{N}_{GT}$	$\hat{N}_{GT}$	$\hat{N}_{GT}$	$\hat{N}_{GT}$
$(\lambda_1,\lambda_2)=(7,5)$	$\hat{N}_{GT}, \hat{N}_{NewLB}$	$\hat{N}_{NewLB}$	$\hat{N}_{NewLB}$	$\hat{N}_{NewLB}$	$\hat{N}_{NewLB}$	$\hat{N}_{NewLB}$
$p_i \sim NB(\kappa,\pi)$						
$\kappa=4,\pi=0.5$	$\hat{N}_{GT}, \hat{N}_{NewLB}$	$\hat{N}_{GT}$	$\hat{N}_{GT}$	$\hat{N}_{GT}$	$\hat{N}_{GT}$	$\hat{N}_{GT}$
$\kappa=4,\pi=0.8$	$\hat{N}_{NewLB}$	$\hat{N}_{NewLB}$	$\hat{N}_{GT}$	$\hat{N}_{GT}$	$\hat{N}_{GT}$	$\hat{N}_{GT}$
$\kappa=9,\pi=0.5$	$\hat{N}_{NewLB}$	$\hat{N}_{NewLB}$	$\hat{N}_{GT}$	$\hat{N}_{GT}$	$\hat{N}_{GT}$	$\hat{N}_{GT}$
$\kappa=9,\pi=0.8$	$\hat{N}_{NewLB}$	$\hat{N}_{GT}, \hat{N}_{NewLB}$	$\hat{N}_{GT}$	$\hat{N}_{GT}$	$\hat{N}_{GT}$	$\hat{N}_{GT}$

Ref. code: 25595809030157UGW

จากตารางที่ 5.3 เมื่อพิจารณาจากค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ ($RMSE$) ภายใต้การแจกแจงปัวซงที่มีพารามิเตอร์ $\lambda = 3,5,7$ ภายใต้การแจกแจงปัวซงแบบผสมที่มีพารามิเตอร์ λ_1, λ_2 ที่มีพารามิเตอร์ $(\lambda_1, \lambda_2) = (3,5)$ และ $(\lambda_1, \lambda_2) = (7,5)$ และ ภายใต้การแจกแจงทวินามเชิงลบ ที่มีพารามิเตอร์ κ, π , $p_i \sim NB(\kappa, \pi)$ โดยที่ $\kappa = (4,9), \pi = (0.5, 0.8)$ สำหรับขนาดประชากรขนาดเล็ก $N = 500{,}1{,}000$ ประชากรขนาดกลาง $N = 5{,}000,\ 10{,}000$ และประชากรขนาดใหญ่ $N = 50{,}000,\ 100{,}000$ สามารถสรุปผลการเปรียบเทียบประสิทธิภาพของตัวประมาณขนาดประชากรทั้ง 4 ตัวได้ดังนี้

ภายใต้ตัวแบบการแจกแจงปัวซงที่มีพารามิเตอร์ $\lambda = 3,5,7$ พบว่า $\hat{N}_{GT}$ และ $\hat{N}_{NewLB}$ ให้ค่า $RMSE$ ต่ำสุดในทุกขนาดประชากรที่ศึกษา

จากผลดังกล่าวข้างต้นสรุปได้ว่า $\hat{N}_{GT}$ และ $\hat{N}_{NewLB}$ มีประสิทธิภาพสูงสำหรับการแจกแจงปัวซงที่มีพารามิเตอร์ $\lambda = 3,5,7$ ในทุกขนาดประชากรที่ศึกษา

ภายใต้ตัวแบบการแจกแจงปัวซงแบบผสม ที่มีพารามิเตอร์ $(\lambda_1, \lambda_2) = (3,5)$ และ $(\lambda_1, \lambda_2) = (7,5)$ สำหรับขนาดประชากร $N = 500{,}1{,}000{,}5{,}000,\ 10{,}000,\ 50{,}000{,}100{,}000$ พบว่า $\hat{N}_{GT}$ และ $\hat{N}_{NewLB}$ ให้ค่า $RMSE$ ต่ำสุด โดยส่วนใหญ่ $\hat{N}_{GT}$ มีค่า $RMSE$ เมื่อพารามิเตอร์ λ มีขนาดเล็ก ในทุกขนาดประชากรที่ศึกษา แต่ $\hat{N}_{NewLB}$ จะมีค่า $RMSE$ ต่ำสุดเมื่อพารามิเตอร์ λ มีค่าสูง ในทุกขนาดประชากรที่ศึกษา

จากผลดังกล่าวข้างต้นสรุปได้ว่า $\hat{N}_{GT}$ และ $\hat{N}_{NewLB}$ มีประสิทธิภาพสูงสำหรับการแจกแจงปัวซงแบบผสม ที่มีพารามิเตอร์ $(\lambda_1, \lambda_2) = (3,5)$ ในทุกขนาดประชากรที่ศึกษา และ $\hat{N}_{NewLB}$ มีประสิทธิภาพสูงสำหรับการแจกแจงปัวซงแบบผสม ที่มีพารามิเตอร์ $(\lambda_1, \lambda_2) = (7,5)$ ในทุกขนาดประชากรที่ศึกษา

ภายใต้ตัวแบบการแจกแจงทวินามเชิงลบ ที่มีพารามิเตอร์ π และ κ โดยที่ $\kappa = (4,9),\ \pi = (0.5, 0.8)$ สำหรับขนาดประชากร $N = 500{,}1{,}000{,}5{,}000,\ 10{,}000{,}50{,}000{,}100{,}000$ พบว่าโดยภาพรวม $\hat{N}_{GT}$ ให้ค่า $RMSE$ ต่ำสุดเกือบทุกขนาดประชากรที่ศึกษา แต่ $\hat{N}_{NewLB}$ มีค่า $RMSE$ ต่ำสุด ในกรณีที่ประชากรมีขนาดเล็ก

จากผลดังกล่าวข้างต้นสรุปได้ว่า $\hat{N}_{NewLB}$ มีประสิทธิภาพสูงสำหรับการแจกแจงทวินามเชิงลบที่มีพารามิเตอร์ π และ κ โดยที่ $\kappa = (4,9),\ \pi = (0.5, 0.8)$ เมื่อประชากรขนาดเล็ก และ $\hat{N}_{GT}$ มีประสิทธิภาพสูงสำหรับการแจกแจงทวินามเชิงลบ ที่มีพารามิเตอร์ π และ κ โดยที่ $\kappa = (4,9),\ \pi = (0.5, 0.8)$ เมื่อประชากรขนาดกลางและขนาดใหญ่

Ref. code: 25595809030157UGW

5.4 ผลการศึกษาจากค่าประมาณสัมประสิทธิ์ความเชื่อมั่นของขนาดประชากร ($1-\hat{\alpha}$)

ตารางที่ 5.4 แสดงผลสรุปช่วงเชื่อมั่นของตัวประมาณขนาดประชากรที่มีประสิทธิภาพสูงสุด เมื่อพิจารณาค่าประมาณสัมประสิทธิ์ความเชื่อมั่นของขนาดประชากร ($1-\hat{\alpha}$)

Model \ N	500	1,000	5,000	10,000	50,000	100,000
$p_i \sim Poi(\lambda)$						
$\lambda=3$	$\hat{N}_{NewLB}$	$\hat{N}_{NewLB}$	$\hat{N}_{NewLB}$	$\hat{N}_{NewLB}$	$\hat{N}_{LB}$ / $\hat{N}_{NewLB}$	$\hat{N}_{NewLB}$
$\lambda=5$	$\hat{N}_{GT}$ / $\hat{N}_{NewLB}$	$\hat{N}_{GT}$ / $\hat{N}_{NewLB}$	$\hat{N}_{GT}$ / $\hat{N}_{NewLB}$	$\hat{N}_{LB}$ / $\hat{N}_{NewLB}$	$\hat{N}_{GT}$ / $\hat{N}_{NewGT}$ $\hat{N}_{NewLB}$	$\hat{N}_{LB}$ / $\hat{N}_{NewLB}$
$\lambda=7$	-	$\hat{N}_{GT}$ / $\hat{N}_{NewGT}$	$\hat{N}_{GT}$ / $\hat{N}_{NewGT}$ / $\hat{N}_{NewLB}$	$\hat{N}_{GT}$ / $\hat{N}_{NewGT}$ / $\hat{N}_{NewLB}$	$\hat{N}_{GT}$ / $\hat{N}_{LB}$ / $\hat{N}_{NewLB}$	$\hat{N}_{NewGT}$ / $\hat{N}_{LB}$
$p_i \sim 0.5Poi(\lambda_1)+0.5Poi(\lambda_2)$						
$(\lambda_1,\lambda_2)=(3,5)$	$\hat{N}_{NewLB}$	$\hat{N}_{NewLB}$	$\hat{N}_{LB}$ / $\hat{N}_{NewLB}$	$\hat{N}_{LB}$	$\hat{N}_{LB}$	-
$(\lambda_1,\lambda_2)=(7,5)$	$\hat{N}_{NewLB}$	$\hat{N}_{NewLB}$	$\hat{N}_{LB}$ / $\hat{N}_{NewLB}$	$\hat{N}_{LB}$ / $\hat{N}_{NewLB}$	$\hat{N}_{NewLB}$	$\hat{N}_{NewLB}$
$p_i \sim NB(\kappa,\pi)$						
$\kappa=4,\pi=0.5$	-	-	-	-	-	-
$\kappa=4,\pi=0.8$	$\hat{N}_{NewLB}$	$\hat{N}_{NewLB}$	$\hat{N}_{NewLB}$	-	-	-
$\kappa=9,\pi=0.5$	-	-	-	$\hat{N}_{GT}$, $\hat{N}_{NewGT}$	-	-
$\kappa=9,\pi=0.8$	$\hat{N}_{NewLB}$	$\hat{N}_{NewLB}$	$\hat{N}_{NewLB}$	-	-	-

จากตารางที่ 5.4 เมื่อพิจารณาจากค่าประมาณสัมประสิทธิ์ความเชื่อมั่นของขนาดประชากร ($1-\hat{\alpha}$) ภายใต้การแจกแจงปัวซงที่มีพารามิเตอร์ $\lambda=3,5,7$ ภายใต้การแจกแจงปัวซงแบบผสมที่มีพารามิเตอร์ λ_1,λ_2 ที่มีพารามิเตอร์ $(\lambda_1,\lambda_2)=(3,5)$ และ $(\lambda_1,\lambda_2)=(7,5)$ และภายใต้การแจกแจงทวินามเชิงลบที่มีพารามิเตอร์ κ,π, $p_i \sim NB(\kappa,\pi)$ โดยที่ $\kappa=(4,9)$, $\pi=(0.5,0.8)$ สำหรับขนาดประชากรขนาดเล็ก $N=500,1,000$ ประชากรขนาดกลาง $N=5,000,\ 10,000$ และประชากรขนาด

ใหญ่ $N = 50,000,\ 100,000$ สามารถสรุปผลการเปรียบเทียบประสิทธิภาพของช่วงเชื่อมั่นของตัวประมาณขนาดประชากรที่สร้างจากตัวประมาณขนาดประชากรทั้ง 4 ตัวได้ดังนี้

ภายใต้ตัวแบบการแจกแจงปัวซงที่มีพารามิเตอร์ $\lambda = 3,5,7$ พบว่าเมื่อพารามิเตอร์ $\lambda = 3$ ช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{NewLB}$ ให้ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นสูงกว่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนดในทุกขนาดประชากรที่ศึกษา

เมื่อพารามิเตอร์ $\lambda = 5$ และประชากรมีขนาดเล็กพบว่า ช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{GT}, \hat{N}_{NewLB}$ ให้ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นสูงกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด และเมื่อประชากรมีขนาดกลางพบว่า ช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{NewGT}, \hat{N}_{NewLB}$ ให้ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นสูงกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด และเมื่อประชากรมีขนาดใหญ่ พบว่าช่วงเชื่อมั่นของขนาดประชากรทั้ง 4 ให้ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นสูงกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด

เมื่อพารามิเตอร์ $\lambda = 7$ พบว่า เมื่อประชากรขนาด $N = 500$ ช่วงเชื่อมั่นของขนาดประชากรทั้ง 4 มีค่าต่ำกว่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด และเมื่อประชากรมีขนาดกลางพบว่า ช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{GT}, \hat{N}_{NewGT}, \hat{N}_{NewLB}$ ให้ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นสูงกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด และเมื่อประชากรมีขนาดใหญ่ พบว่าช่วงเชื่อมั่นของขนาดประชากรทั้ง 4 คือ $\hat{N}_{GT}, \hat{N}_{LB}, \hat{N}_{NewGT}$ และ $\hat{N}_{NewLB}$ ให้ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นสูงกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด

ภายใต้ตัวแบบการแจกแจงปัวซงแบบผสมที่มีพารามิเตอร์ λ_1, λ_2 ที่มีพารามิเตอร์ $(\lambda_1, \lambda_2) = (3,5)$ และ $(\lambda_1, \lambda_2) = (7,5)$ สำหรับ $N = 500, 1,000, 5,000, 10,000, 50,000, 100,000$ พิจารณาเมื่อพารามิเตอร์ $(\lambda_1, \lambda_2) = (3,5)$ พบว่า เมื่อประชากรมีขนาดเล็ก ช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{NewLB}$ ให้ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นสูงกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด และเมื่อประชากรมีขนาดใหญ่ขึ้นพบว่า ช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{LB}$ ให้ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นสูงกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด ยกเว้นเมื่อ $N = 100,000$ ที่ช่วงเชื่อมั่นของขนาดประชากรทั้ง 4 ให้ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด

เมื่อพารามิเตอร์ $(\lambda_1, \lambda_2) = (7,5)$ พบว่า ช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{NewLB}$ ให้ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นสูงกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนดในทุกขนาดประชากรที่ศึกษา นอกจากนี้เมื่อประชากรมีขนาดกลางยังพบว่า ช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{LB}$ ให้ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นสูงกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด

ภายใต้ตัวแบบการแจกแจงทวินามเชิงลบ ที่มีพารามิเตอร์ π และ κ โดยที่ $\kappa=(4,9), \pi=(0.5,0.8)$ สำหรับขนาดประชากร $N=500,1,000,5,000, 10,000,50,000,100,000$ พบว่าเมื่อพารามิเตอร์ $\kappa=4, \pi=0.5$ ช่วงเชื่อมั่นของขนาดประชากรทั้ง 4 ให้ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนดในทุกขนาดประชากรที่ศึกษา

เมื่อพารามิเตอร์ $\kappa=9, \pi=0.5$ พบว่า เมื่อประชากร $N=10,000$ พบว่าช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{GT}, \hat{N}_{NewGT}$ ให้ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นสูงกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด

เมื่อพารามิเตอร์ $\kappa=4, \pi=0.8$ และ $\kappa=9, \pi=0.8$ เมื่อประชากรมีขนาด $N=500,1,000,5,000$ พบว่า ช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{NewLB}$ ให้ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นสูงกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด

5.5 ผลการศึกษาจากค่าประมาณความกว้างเฉลี่ยของช่วงเชื่อมั่น (Average Length)

ตารางที่ 5.5 แสดงผลสรุปช่วงเชื่อมั่นของตัวประมาณขนาดประชากรที่มีประสิทธิภาพสูงสุดเมื่อพิจารณาค่าประมาณความกว้างเฉลี่ยของช่วงเชื่อมั่น (Average Length)

Model \ N	500	1,000	5,000	10,000	50,000	100,000
$p_i \sim Poi(\lambda)$						
$\lambda=3$	$\hat{N}_{NewLB}$	$\hat{N}_{NewLB}$	$\hat{N}_{NewLB}$	$\hat{N}_{NewLB}$	$\hat{N}_{NewLB}$	$\hat{N}_{NewLB}$
$\lambda=5$	$\hat{N}_{GT}$	$\hat{N}_{GT}$	$\hat{N}_{NewGT}$	$\hat{N}_{NewLB}$	$\hat{N}_{NewGT}$	$\hat{N}_{NewLB}$
$\lambda=7$	-	$\hat{N}_{NewGT}$	$\hat{N}_{NewLB}$	$\hat{N}_{NewLB}$	$\hat{N}_{GT}$	$\hat{N}_{NewGT}$
$p_i \sim 0.5Poi(\lambda_1)+0.5Poi(\lambda_2)$						
$(\lambda_1,\lambda_2)=(3,5)$	$\hat{N}_{NewLB}$	$\hat{N}_{NewLB}$	$\hat{N}_{NewLB}$	$\hat{N}_{LB}$	$\hat{N}_{LB}$	-
$(\lambda_1,\lambda_2)=(7,5)$	$\hat{N}_{NewLB}$	$\hat{N}_{NewLB}$	$\hat{N}_{NewLB}$	$\hat{N}_{NewLB}$	$\hat{N}_{NewLB}$	$\hat{N}_{NewLB}$
$p_i \sim NB(\kappa,\pi)$						
$\kappa=4, \pi=0.5$	-	-	-	-	-	-
$\kappa=4, \pi=0.8$	$\hat{N}_{NewLB}$	$\hat{N}_{NewLB}$	$\hat{N}_{NewLB}$	-	-	-
$\kappa=9, \pi=0.5$	-	-	-	$\hat{N}_{GT}$	-	-
$\kappa=9, \pi=0.8$	$\hat{N}_{NewLB}$	$\hat{N}_{NewLB}$	$\hat{N}_{NewLB}$	-	-	-

จากตารางที่ 5.5 เมื่อพิจารณาจากค่าประมาณความกว้างเฉลี่ยของช่วงเชื่อมั่น (Average Length) ภายใต้การแจกแจงปัวซงที่มีพารามิเตอร์ $\lambda = 3,5,7$ ภายใต้การแจกแจงปัวซงแบบผสมที่มีพารามิเตอร์ λ_1, λ_2 ที่มีพารามิเตอร์ $(\lambda_1, \lambda_2) = (3,5)$ และ $(\lambda_1, \lambda_2) = (7,5)$ และภายใต้การแจกแจงทวินามเชิงลบที่มีพารามิเตอร์ κ, π, $p_i \sim NB(\kappa, \pi)$ โดยที่ $\kappa = (4,9), \pi = (0.5, 0.8)$ สำหรับขนาดประชากรขนาดเล็ก $N = 500, 1,000$ ประชากรขนาดกลาง N =5,000, 10,000 และประชากรขนาดใหญ่ N =50,000, 100,000 สามารถสรุปผลการเปรียบเทียบประสิทธิภาพของช่วงเชื่อมั่นของตัวประมาณขนาดประชากรที่สร้างจากตัวประมาณขนาดประชากรทั้ง 4 ตัวได้ดังนี้

ภายใต้ตัวแบบการแจกแจงปัวซงที่มีพารามิเตอร์ $\lambda = 3,5,7$ พิจารณาเมื่อ $\lambda = 3$ พบว่าช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{NewLB}$ มีค่าประมาณความกว้างเฉลี่ยของช่วงเชื่อมั่นที่แคบที่สุดในทุกๆ ขนาดประชากร สำหรับช่วงเชื่อมั่นของขนาดประชากรอื่นๆ ไม่พิจารณา เนื่องจากไม่มีช่วงเชื่อมั่นของขนาดประชากรใดที่ให้ค่าสูงกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด

เมื่อพิจารณา $\lambda = 5$ เมื่อประชากรขนาดเล็ก พบว่าช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{GT}$ มีค่าประมาณความกว้างเฉลี่ยของช่วงเชื่อมั่นที่แคบที่สุด และเมื่อประชากรมีขนาดกลาง พบว่าช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{NewGT}$ มีค่าประมาณความกว้างเฉลี่ยของช่วงเชื่อมั่นที่แคบที่สุด และเมื่อประชากมีขนาดใหญ่ พบว่าช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{NewLB}$ มีค่าประมาณความกว้างเฉลี่ยของช่วงเชื่อมั่นที่แคบที่สุด

เมื่อพิจารณา $\lambda = 7$ เมื่อประชากรมีขนาดเล็ก พบว่าช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{NewGT}$ มีค่าประมาณความกว้างเฉลี่ยของช่วงเชื่อมั่นที่แคบที่สุด และเมื่อประชากรขนาดกลางพบว่าช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{NewLB}$ มีค่าประมาณความกว้างเฉลี่ยของช่วงเชื่อมั่นที่แคบที่สุด และเมื่อประชากรมีขนาดใหญ่ $N = 100,000$ พบว่าช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{NewGT}$ มีค่าประมาณความกว้างเฉลี่ยของช่วงเชื่อมั่นที่แคบที่สุด

ภายใต้ตัวแบบการแจกแจงปัวซงแบบผสมที่มีพารามิเตอร์ λ_1, λ_2 ที่มีพารามิเตอร์ $(\lambda_1, \lambda_2) = (3,5)$ และ $(\lambda_1, \lambda_2) = (7,5)$ สำหรับ $N = 500, 1,000, 5,000, 10,000, 50,000, 100,000$ พิจารณาเมื่อ $(\lambda_1, \lambda_2) = (3,5)$ ช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{NewLB}$ มีค่าประมาณความกว้างเฉลี่ยของช่วงเชื่อมั่นที่แคบที่สุดเมื่อประชากรมีขนาดเล็กและขนาดกลาง และเมื่อประชากรมีขนาดใหญ่ ช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{LB}$ จะมีค่าประมาณความกว้างเฉลี่ยของช่วงเชื่อมั่นที่แคบที่สุด

Ref. code: 25595809030157UGW

พิจารณาเมื่อ $(\lambda_1,\lambda_2)=(7,5)$ พบว่าช่วงเชื่อมั่นของขนาดประชากรโดย $\hat{N}_{NewLB}$ มีค่าประมาณความกว้างเฉลี่ยของช่วงเชื่อมั่นที่แคบที่สุดในทุกขนาดประชากรที่ศึกษา

เมื่อพิจารณาภายใต้ตัวแบบการแจกแจงทวินามเชิงลบ ที่มีพารามิเตอร์ π และ κ โดยที่ $\kappa=(4,9),\pi=(0.5,0.8)$ สำหรับขนาดประชากร $N=500,1{,}000,5{,}000, 10{,}000,50{,}000,100{,}000$ พบว่า กรณีที่ $\kappa=4,\pi=0.5$ ไม่พิจารณาเนื่องจากไม่มีช่วงเชื่อมั่นของขนาดประชากรใดที่ให้ค่าสูงกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด

5.6 ผลการศึกษาเปรียบเทียบประสิทธิภาพของตัวประมาณขนาดประชากร

ผลการศึกษาสรุปได้ว่า ตัวประมาณขนาดประชากรทั้ง 4 ตัวคือ $\hat{N}_{GT},\hat{N}_{NewGT},\hat{N}_{LB}$ และ $\hat{N}_{NewLB}$ มีประสิทธิภาพแตกต่างกัน ตามรายละเอียดดังต่อไปนี้

$\hat{N}_{NewGT}$ เป็นตัวประมาณขนาดประชากรที่ปรับปรุงโดยใช้การถดถอยควอนไทล์ที่ระดับ 0.50 ผู้วิจัยได้จำลองสถานการณ์ภายใต้ข้อมูลลักษณะ Heterogeneous poisson model หรือการแจกแจงปัวซงที่มีพารามิเตอร์ λ พบว่า เมื่อเปรียบเทียบประสิทธิภาพกับ $\hat{N}_{GT}$ ที่เป็นตัวประมาณขนาดประชากรที่อาศัยการถดถอยเชิงเส้นแบบถ่วงน้ำหนักนั้น การใช้การถดถอยควอนไทล์ที่ระดับ 0.50 มาประมาณค่า κ ไม่ได้ปรับปรุงประสิทธิภาพตัวประมาณให้ดีขึ้น เมื่อข้อมูลมีแจกแจงปัวซงที่มีพารามิเตอร์ λ

เมื่อจำลองสถานการณ์ภายใต้ข้อมูลลักษณะ Heterogeneous poisson model ได้แก่ การแจกแจงปัวซงแบบผสมที่มีพารามิเตอร์ λ_1,λ_2 การใช้การถดถอยควอนไทล์ที่ระดับ 0.50 ไม่ได้ปรับปรุงประสิทธิภาพตัวประมาณให้ดีขึ้น เมื่อข้อมูลมีแจกแจงปัวซงแบบผสมที่มีพารามิเตอร์ λ_1,λ_2

เมื่อจำลองสถานการณ์ภายใต้การแจกแจงทวินามเชิงลบที่มีพารามิเตอร์ κ,π พบว่า $\hat{N}_{NewGT}$ มีประสิทธิภาพสูงกว่า $\hat{N}_{GT}$ ในแง่ของค่าความเอนเอียงสัมพัทธ์ ($RBias$) เมื่อพารามิเตอร์ $\kappa=4,\pi=0.5$ สอดคล้องกับกับงานวิจัยของ Pijitrattana & Lerdsuwansri ในปี ค.ศ. 2015 (Pijitrattana&Lerdsuwansri, 2015) ซึ่ง Pijitrattana & Lerdsuwansri พบว่า $\hat{N}_{GT}$ จะมีประสิทธิภาพต่ำเมื่อพารามิเตอร์ κ ขนาดเล็ก เนื่องจากค่าของ κ ที่ประมาณได้จะมีค่าห่างไกลจากค่าจริง หรือทำให้มีค่าความเอนเอียงสัมพัทธ์สูง แต่จะมีค่าความแปรปรวนสัมพัทธ์ต่ำ

เมื่อพิจารณาการแจกแจงทวินามเชิงลบที่มีพารามิเตอร์ $\kappa=4,\ \pi=0.8$, $\kappa=9,\ \pi=0.5$ และ $\kappa=9,\ \pi=0.8$ พบว่า การใช้การถดถอยควอนไทล์ที่ระดับ 0.50 มาประมาณค่า κ สามารถปรับปรุงประสิทธิภาพได้ในแง่ของความเอนเอียงสัมพัทธ์เมื่อประชากรมีขนาดกลางและขนาดใหญ่

ต่อมาพิจารณาเปรียบเทียบประสิทธิภาพของ $\hat{N}_{LB}$ กับ $\hat{N}_{NewLB}$ ผู้วิจัยได้จำลองสถานการณ์ภายใต้ข้อมูลลักษณะ Heterogeneous poisson model หรือการแจกแจงปัวซงที่มีพารามิเตอร์ λ พบว่า $\hat{N}_{NewLB}$ มีประสิทธิภาพสูงทุกขนาดประชากรที่ศึกษา เนื่องจากผู้วิจัยได้จำลองภายใต้ค่าเฉลี่ยการปรากฏของข้อมูลมีค่าสูงหรืออยู่ปลายหาง ทำให้เอื้อต่อการใช้การประมาณขนาดประชากรที่ใช้การถดถอยควอนไทล์ที่ระดับ 0.50 ($\hat{N}_{NewLB}$) สอดคล้องกับงานวิจัยของ Lanumteang&Böhning ในปี ค.ศ. 2011 (Lanumteang&Böhning, 2011) ซึ่ง Lanumteang&Böhning พบว่า $\hat{N}_{LB}$ ให้ค่าความเอนเอียงต่ำ แต่มีค่าความแปรปรวนสูง และจากสูตรการหาตัวประมาณขนาดประชากรของ $\hat{N}_{LB}$ ได้อาศัยค่า f_1, f_2 และ f_3 จากการจำลองสถานการณ์ทำให้ค่า f_1, f_2 และ f_3 มีค่าน้อย จึงทำให้ผลสรุปที่ได้นั้นพบว่า $\hat{N}_{NewLB}$ มีประสิทธิภาพสูงกว่า $\hat{N}_{LB}$

เมื่อจำลองสถานการณ์ภายใต้ข้อมูลลักษณะ Heterogeneous poisson model ได้แก่ การแจกแจงปัวซงแบบผสมที่มีพารามิเตอร์ จะเห็นได้ชัดเมื่อ จำลองภายใต้ข้อมูลที่มีพารามิเตอร์ $(\lambda_1, \lambda_2) = (7,5)$ พบว่า $\hat{N}_{NewLB}$ มีประสิทธิภาพสูงกว่า $\hat{N}_{LB}$

เมื่อจำลองสถานการณ์ภายใต้การแจกแจงทวินามเชิงลบที่มีพารามิเตอร์ κ, π พบว่า $\hat{N}_{LB}$ มีประสิทธิภาพสูงในแง่ค่าความเอนเอียงสัมพัทธ์ แต่ $\hat{N}_{NewLB}$ มีประสิทธิภาพสูงในแง่ของค่าความแปรปรวนและค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ในทุกขนาดประชากรที่ศึกษา สอดคล้องกับงานวิจัยของ Lanumteang&Böhning ในปี ค.ศ. 2011 (Lanumteang&Böhning, 2011) ซึ่ง Lanumteang&Böhning พบว่า $\hat{N}_{LB}$ ให้ค่าความเอนเอียงต่ำ แต่มีค่าความแปรปรวนสูง

5.7 ผลการศึกษาเปรียบเทียบประสิทธิภาพของช่วงเชื่อมั่นของขนาดประชากร

ผลการศึกษาสรุปได้ว่า ช่วงเชื่อมั่นของขนาดประชากรที่สร้างจากตัวประมาณขนาดประชากรทั้ง 4 ตัวคือ $\hat{N}_{GT}, \hat{N}_{NewGT}, \hat{N}_{LB}$ และ $\hat{N}_{NewLB}$ พบว่าช่วงเชื่อมั่นที่สร้างจาก $\hat{N}_{NewLB}$ มีประสิทธิภาพเมื่อข้อมูลมีการแจกแจงปัวซงที่มีพารามิเตอร์ $\lambda = 3$ ในทุกขนาดประชากรที่ศึกษา

เมื่อข้อมูลมีการแจกแจงปัวซงแบบผสมที่มีพารามิเตอร์ $(\lambda_1, \lambda_2) = (3,5)$ ช่วงเชื่อมั่นที่สร้างจาก $\hat{N}_{NewLB}$ มีประสิทธิภาพเมื่อประชากรมีขนาดเล็ก และเมื่อ $(\lambda_1, \lambda_2) = (7,5)$ ช่วงเชื่อมั่นที่สร้างจาก $\hat{N}_{NewLB}$ มีประสิทธิภาพทุกขนาดประชากรที่ศึกษา

และเมื่อข้อมูลมีการแจกแจงทวินามเชิงลบที่มีพารามิเตอร์ κ, π ที่ $\kappa = 4, \pi = 0.8$ และ $\kappa = 9, \pi = 0.8$ พบว่า ช่วงเชื่อมั่นที่สร้างจาก $\hat{N}_{NewLB}$ มีประสิทธิภาพเมื่อประชากรมีขนาดเล็ก

Ref. code: 25595809030157UGW

5.8 ข้อเสนอแนะ

5.8.1 ข้อเสนอแนะเกี่ยวกับการวิจัย

1. จากการจำลองสถานการณ์เป็นการจำลองภายใต้ลักษณะข้อมูลที่เอื้อต่อการประมาณค่าโดยใช้การถดถอยควอนไทล์ที่ระดับ 0.50 ดังนั้นจึงควรจำลองภายใต้สถานการณ์ที่แตกต่างกัน เพื่อพิจารณาประสิทธิภาพของตัวประมาณที่ได้พัฒนาขึ้นนั้นได้อย่างครอบคลุม

2. ในงานวิจัยนี้ มีข้อจำกัดที่ว่าค่าสัมประสิทธิ์การถดถอย α ต้องมีค่ามากกว่า 0 และจากการหาค่าประมาณความแปรปรวนของ $\hat{N}_{NewLB}$ นั้น ผู้วิจัยกำหนดให้ค่าสัมประสิทธิ์การถดถอย α, β เป็นอิสระต่อกัน นั่นคือไม่ได้บวกเทอม $2Cov(\alpha, \beta)$ ทำให้ค่าประมาณความแปรปรวนของ $\hat{N}_{NewLB}$ มีค่าน้อย

5.8.2 ข้อเสนอแนะเกี่ยวกับการนำไปประยุกต์ใช้

1. จากการได้จำลองสถานการณ์ภายใต้การแจกแจงเรขาคณิต (Geometric distribution) เพื่อเปรียบเทียบประสิทธิภาพของตัวประมาณและเปรียบเทียบประสิทธิภาพของช่วงเชื่อมั่นของขนาดประชากร พบว่า ค่าที่ได้ของตัวประมาณขนาดประชากรทั้ง 4 ไม่เหมาะสมในทุกๆ สถานการณ์ เนื่องจาก การแจกแจงทวินามเชิงลบ (Negative Binomial distribution) ที่มีพารามิเตอร์ κ, π เมื่อ $\kappa = 1$ จะเป็นการแจกแจงเรขาคณิต ซึ่งสมการถดถอยเชิงเส้นภายใต้การแจกแจงเรขาคณิตจะผ่านจุด $(0,0)$ ซึ่งทำให้พารามิเตอร์ α ในตัวแบบมีค่าเท่ากับ 0 ซึ่งไม่ใช่สมการถดถอยเชิงเส้นที่ใช้ในการประมาณค่าของขนาดประชากรที่ศึกษา ด้วยเหตุนี้ จึงทำให้ค่าประมาณที่ได้ภายใต้การแจกแจงเรขาคณิตไม่เหมาะสมในทุกๆ สถานการณ์

Ref. code: 25595809030157UGW

รายการอ้างอิง

หนังสือและบทความในหนังสือ

A.Colin Cameron and Pravin K. Trivedi, Microeconometrics Methods and Applications, Cambridge University.

บทความวารสาร

Bohning,D., Dietz, E., Kuhnert, R., Schon, D. (2005),“Mixture models for Capture-Recapture count data,” Statistical Methods & Applications (2005) 14:29-43.

Bohning,D.(2008),“A Simple Variance Formula for Population Size Estimators by Conditioning,” Statistical Methodology,5,410-423.

Bohning,D., Baksh, M.F.,Lerdsuwansri, R.,and Gallagher,J.(2013),“Use of the Ratio Plot in capture-Recapture Estimation,” Journal of Computational and Graphical Statistics,Volume22,Number 1,135-155.

Buchinsky, M. (1998),“Recent Advances in Quantile Regression Models,” The Journal of Human Resources, Vol.33, No.1 (Winter, 1998), pp.88-126.

Chao,A.(1987),“Estimating the Population Size for Capture-Recapture Data With Unequal Catchability,” Biometrics, 43,783-791.

Chao,A.,and Huggins,R.M.(2005),“Modern Closed-Population Capture-Recapture Model,” in Handbook of Capture-Recapture, eds.S.C.Amstrup,T.L.Mcdonald, and B.F.J. Manly,Princeton: Princeton University Press,58-87.

Lanumtaeng,K., Bohning,D.(2011),“An extension of Chao’s estimator of population size based on the first three capture frequency counts,” Computational Statistics and Data Analysis,55,2302-2311.

Vand der Heijen, P . G. M., Bustami, R., Cruyff, M., Engbersen, G., and van Houwelingen, H. C. (2003a),“Point and Interval Estimation of the Population Size Using the Truncated Poisson regression Model,” Statistical Modelling, 3, 305-322.

Ref. code: 25595809030157UGW

Zelterman, D.(1988),“Robust Estimation in Truncated Discrete Distributions With Applications to Capture-Recapture Experiments,” Journal of Statistical Planning and Inference,18,225-237.

กัญญ์สุดา นิ่มอนุสสรณ์กุล และคณะ (2555),“การวิเคราะห์ผลกระทบของหนี้สาธารณะ ต่ออัตราการเจริญเติบโตทางเศรษฐกิจ ด้วยวิธีการถดถอยแบบควอนไทล์,” CMU. JOURNAL OF ECONOMICS.

ศราวุฒิ ชุมภูอินตา (2558), “ปัจจัยทางเศรษฐกิจมหภาคที่มีผลต่ออัตราผลตอบแทนของดัชนีราคาหุ้นกลุ่มพลังงานโดยวิธีการถดถอยควอนไทล์” การประชุมวิชาการระดับชาติ มหาวิทยาลัยราชภัฏสวนสุนันทา.

วิทยานิพนธ์

Pijitrattana, P.(2014). Confidence Interval for Population size Using the Generalized Turing Estimator. M.S. (Appied Statistics) Thesis, Thammasat University.

Ref. code: 25595809030157UGW

ภาคผนวก

Ref. code: 25595809030157UGW

ภาคผนวก ก
โปรแกรมที่ใช้ในการศึกษา

1. เปรียบเทียบประสิทธิภาพของตัวประมาณขนาดประชากร

การศึกษาใช้การจำลองข้อมูลและประมวลผล โดยการเขียนคำสั่งในโปรแกรม R Studio โดยมีคำสั่ง แสดงได้ดังนี้

```
######################################################################
##  For generate random number of Poisson distribution, Geometric distribution and
Negative Binomial distribution    ##
######################################################################

        library(quantreg)
        getdata<-function(getx,getfx)   ## adjust for freq = 0
        {
          xlist <- c(0:max(getx))
          if(length(getx)==length(xlist)) { outx <-
matrix(cbind(getx,getfx),nrow=length(getx),ncol=2,byrow=FALSE)}
          else {
            "%w/o%" <- function(u,v) u[!u %in% v] #--  u without v
            xmis<- xlist %w/o% getx
            fmis <- rep(NA,length(xmis))
            mis_fig<-cbind(xmis,fmis)
            obs_fig<-cbind(getx,getfx)
            com_list<-rbind(obs_fig,mis_fig)
            aa<-com_list[,1]
            bb<-com_list[,2]
            o<-order(aa)
            outx <-
matrix(cbind(aa[o],bb[o]),nrow=length(xlist),ncol=2,byrow=FALSE)}
```

```
  return(outx)
}

### main program ##
est.GTsimple<-function(B,N,dist,para)
{
  khat<-rep(0,B)
  Nhat.GT <-rep(0,B)
  VarNhat.GT <-rep(0,B)

  Rbias.GT=rep(0,B)
  RVar.GT=rep(0,B)
  RMSE.GT=rep(0,B)

  i<- 1
  set.seed(12345)
  while(i<=B)
  {
    if (dist=="pois") {x <- rpois(N,para)}
    else if (dist=="geom") {x <- rgeom(N,para)}
    else if (dist=="nbinom") {x <- rnbinom(N,para[1],para[2])}
    s <- sum(x)
    f1 <-sum(as.numeric(x==1))
    f2 <- sum(as.numeric(x==2))
    f3 <- sum(as.numeric(x==3))
    tab<-table(x)
    dtab <- as.data.frame(tab)
    out<-attributes(tab)
    outx<-out$dimnames$x
    getx<-as.numeric(outx)
    getfx<- as.numeric(dtab$Freq)  ## get fx
```

```
size<-max(getx)
xdat <- getdata(getx,getfx)
obsdat <- xdat[-1,]  ### observed frequency
n <- sum(obsdat[,2],na.rm=TRUE)
newsize = size
track.na = 0
for(j in 1:size){
  if(is.na(obsdat[j,2]))
  {track.na = cbind(track.na,j)}
}
track.na = track.na[-1]
if(length(track.na > 0)){
  obsdat = obsdat[-track.na,]
  size = size-length(track.na)}
r_x <-rep(0,size)
w_x <-rep(0,size)

for (j in 1:size)
{
  if(j==size){r_x[j]<-NA; w_x[j]<-NA }
  else{r_x[j]<-obsdat[j+1,1]*obsdat[j+1,2]/obsdat[j,2]
  w_x[j]<-((r_x[j])^2*((1/obsdat[j+1,2])+(1/obsdat[j,2])))^(-1)}
} ## end loop j

ex_na <- na.omit(r_x)
pnt_no <- length(ex_na)
if(pnt_no > 2){
  mod1 <- lm(r_x~obsdat[,1],weights=w_x)
  sum <- summary(mod1)
  alpha_hat1 <- sum$coefficients[1]
  beta_hat1 <- sum$coefficients[2]
```

Ref. code: 25595809030157UGW

```
if((alpha_hat1>0)&&(beta_hat1>0)&&(beta_hat1<1))
{
  #GT#
  khat[i] <- alpha_hat1/beta_hat1
  p0 <- (f1/s)^(khat[i]/(khat[i]+1))
  Nhat.GT[i] <- n/(1-p0)
  Var1<-(n*p0)/((1-p0)^2)
  A1 <- (n^2)/((1-p0)^4)
  A2 <- s^((-2*khat[i])/(khat[i]+1))
  A3 <- (khat[i]/(khat[i]+1))^2
  A4 <- f1^((khat[i]-1)/(khat[i]+1))
  A5 <- 1-(f1/Nhat.GT[i])
  A6 <- s^((-3*khat[i])/(khat[i]+1))
  A7 <- f1^(((2*khat[i])-1)/(khat[i]+1))
  Var2<-  A1*((A2*A3*A4*A5)+(A6*A3*A7))
  VarNhat.GT[i] <- Var1+Var2
  i <- i+1}
}
else{ i <- i}
}##end B replicates
mean_khat <- mean(khat)
meanNGT <- mean(Nhat.GT)
Rbias.GT=(meanNGT-N)/N
RVar.GT=mean((Nhat.GT-meanNGT)^2,na.rm=TRUE)/N^2
RMSE.GT=mean((Nhat.GT-N)^2,na.rm=TRUE)/N^2
VarNhatGT=mean(VarNhat.GT)
seNhatGT=sqrt(VarNhatGT)
outcome<-cbind(meanNGT,Rbias.GT,RVar.GT,RMSE.GT,VarNhatGT,seNhatGT)
return(outcome)
} ##end program
```

Ref. code: 25595809030157UGW

```
est.GTquantile<-function(B,N,dist,para)
{
  khat<-rep(0,B)
  Nhat.newGT <-rep(0,B)
  VarNhat.newGT <-rep(0,B)
  Rbias.newGT=rep(0,B)
  RVar.newGT=rep(0,B)
  RMSE.newGT=rep(0,B)

  i<- 1
  set.seed(12345)
  while(i<=B)
  {
    if (dist=="pois") {x <- rpois(N,para)}
    else if (dist=="geom") {x <- rgeom(N,para)}
    else if (dist=="nbinom") {x <- rnbinom(N,para[1],para[2])}
    s <- sum(x)
    f1 <-sum(as.numeric(x==1))
    f2 <- sum(as.numeric(x==2))
    f3 <- sum(as.numeric(x==3))
    tab<-table(x)
    dtab <- as.data.frame(tab)
    out<-attributes(tab)
    outx<-out$dimnames$x
    getx<-as.numeric(outx)
    getfx<- as.numeric(dtab$Freq)  ## get fx
    size<-max(getx)
    xdat <- getdata(getx,getfx)
    obsdat <- xdat[-1,]  ### observed frequency
    n <- sum(obsdat[,2],na.rm=TRUE)
    newsize = size
```

```
track.na = 0
for(j in 1:size){
  if(is.na(obsdat[j,2]))
  {track.na = cbind(track.na,j)}
}
track.na = track.na[-1]
if(length(track.na > 0)){
  obsdat = obsdat[-track.na,]
  size = size-length(track.na)}
r_x <-rep(0,size)
w_x <-rep(0,size)

for (j in 1:size)
{
  if(j==size){r_x[j]<-NA; w_x[j]<-NA }
  else{r_x[j]<-obsdat[j+1,1]*obsdat[j+1,2]/obsdat[j,2]
  w_x[j]<-((r_x[j])^2*((1/obsdat[j+1,2])+(1/obsdat[j,2])))^(-1)}
} ## end loop j

ex_na <- na.omit(r_x)
pnt_no <- length(ex_na)
if(pnt_no > 2){ mod3 <- rq(r_x~obsdat[,1],weights=w_x,na.action =
na.omit)
sum <- summary.rq(mod3)
alpha_hat3<- sum$coefficients[1]
beta_hat3 <- sum$coefficients[2]
if((alpha_hat3>0)&&(beta_hat3>0)&&(beta_hat3<1))

{

  #newGT#
```

```
khat[i] <- alpha_hat3/beta_hat3
p0 <- (f1/s)^(khat[i]/(khat[i]+1))
Nhat.newGT[i] <- n/(1-p0)
Var1<-(n*p0)/((1-p0)^2)
A1 <- (n^2)/((1-p0)^4)
A2 <- s^((-2*khat[i])/(khat[i]+1))
A3 <- (khat[i]/(khat[i]+1))^2
A4 <- f1^((khat[i]-1)/(khat[i]+1))
A5 <- 1-(f1/Nhat.newGT[i])
A6 <- s^((-3*khat[i])/(khat[i]+1))
A7 <- f1^(((2*khat[i])-1)/(khat[i]+1))
Var2<-  A1*((A2*A3*A4*A5)+(A6*A3*A7))
VarNhat.newGT[i] <- Var1+Var2
i <- i+1}
}
else{ i <- i}
}##end B replicates
mean_khat <- mean(khat)
meannewGT <- mean(Nhat.newGT)
Rbias.newGT=(meannewGT-N)/N
RVar.newGT=mean((Nhat.newGT-meannewGT)^2,na.rm=TRUE)/N^2
RMSE.newGT=mean((Nhat.newGT-N)^2,na.rm=TRUE)/N^2
VarNhatnewGT=mean(VarNhat.newGT)
seNhatnewGT=sqrt(VarNhatnewGT)
outcome<-cbind(meannewGT,Rbias.newGT,RVar.newGT,RMSE.newGT,
VarNhatnewGT,seNhatnewGT)
return(outcome)
} ##end program

est.LBsimple<-function(B,N,dist,para)
{
```

```
Nhat.LB <-rep(0,B)
VarNhat.LB <-rep(0,B)
Rbias.LB=rep(0,B)
RVar.LB=rep(0,B)
RMSE.LB=rep(0,B)

i<- 1
set.seed(12345)
while(i<=B)
{
  if (dist=="pois") {x <- rpois(N,para)}
  else if (dist=="geom") {x <- rgeom(N,para)}
  else if (dist=="nbinom") {x <- rnbinom(N,para[1],para[2])}
  s <- sum(x)
  f1 <-sum(as.numeric(x==1))
  f2 <- sum(as.numeric(x==2))
  f3 <- sum(as.numeric(x==3))
  tab<-table(x)
  dtab <- as.data.frame(tab)
  out<-attributes(tab)
  outx<-out$dimnames$x
  getx<-as.numeric(outx)
  getfx<- as.numeric(dtab$Freq)  ## get fx
  size<-max(getx)
  xdat <- getdata(getx,getfx)
  obsdat <- xdat[-1,]  ### observed frequency
  n <- sum(obsdat[,2],na.rm=TRUE)
  newsize = size
  track.na = 0
  for(j in 1:size){
    if(is.na(obsdat[j,2]))
```

```
  {track.na = cbind(track.na,j)}
}
track.na = track.na[-1]
if(length(track.na > 0)){
  obsdat = obsdat[-track.na,]
  size = size-length(track.na)}
r_x <-rep(0,size)
w_x <-rep(0,size)

for (j in 1:size)
{
  if(j==size){r_x[j]<-NA; w_x[j]<-NA }
  else{r_x[j]<-obsdat[j+1,1]*obsdat[j+1,2]/obsdat[j,2]
  w_x[j]<-((r_x[j])^2*((1/obsdat[j+1,2])+(1/obsdat[j,2])))^(-1)}
} ## end loop j

ex_na <- na.omit(r_x)
pnt_no <- length(ex_na)
if(pnt_no > 2){ mod1 <- lm(log(r_x)~(obsdat[,1]+1),weights=w_x)
sum <- summary(mod1)
alpha_hat1<- sum$coefficients[1]
beta_hat1 <- sum$coefficients[2]
if((alpha_hat1>-1)&&(beta_hat1>-Inf)&&(beta_hat1<Inf))
{
  #LB#

  Nhat.LB[i]<- n+(3*f3*f1^3)/(4*f2^3)
  LB1<-(9/4)^2*(f1^5*f3^2/f2^6)*(f1/f2+1)
  LB2<-(3/4)^2*(f1^6*f3/f2^6)*(1-f3/n)
  LB3<-(3*n/4)*(f1^3*f3)/(n*f2^3+3*f1^3*f3/4)
  VarNhat.LB[i]<-LB1+LB2+LB3
```

Ref. code: 25595809030157UGW

```
      i <- i+1}
    }
    else{ i <- i}
  }##end B replicates
  meanNLB <- mean(Nhat.LB)
  Rbias.LB=(meanNLB-N)/N
  RVar.LB=mean((Nhat.LB-meanNLB)^2,na.rm=TRUE)/N^2
  RMSE.LB=mean((Nhat.LB-N)^2,na.rm=TRUE)/N^2
  varNhatLB=mean(VarNhat.LB)
  seNhatLB=sqrt(varNhatLB)

 outcome<-cbind(meanNLB,Rbias.LB,RVar.LB,RMSE.LB,varNhatLB,seNhatLB)
  return(outcome)
} ##end program

est.LBquantile<-function(B,N,dist,para)
{
  A<-rep(0,B)
  Nhat.newLB <-rep(0,B)
  VarNhat.newLB <-rep(0,B)
  Rbias.newLB=rep(0,B)
  RVar.newLB=rep(0,B)
  RMSE.newLB=rep(0,B)

  i<- 1
  set.seed(12345)
  while(i<=B)
  {
    if (dist=="pois") {x <- rpois(N,para)}
    else if (dist=="geom") {x <- rgeom(N,para)}
    else if (dist=="nbinom") {x <- rnbinom(N,para[1],para[2])}
```

```
s <- sum(x)
f1 <-sum(as.numeric(x==1))
f2 <- sum(as.numeric(x==2))
f3 <- sum(as.numeric(x==3))
tab<-table(x)
dtab <- as.data.frame(tab)
out<-attributes(tab)
outx<-out$dimnames$x
getx<-as.numeric(outx)
getfx<- as.numeric(dtab$Freq)  ## get fx
size<-max(getx)
xdat <- getdata(getx,getfx)
obsdat <- xdat[-1,]  ### observed frequency
n <- sum(obsdat[,2],na.rm=TRUE)
newsize = size
track.na = 0
for(j in 1:size){
  if(is.na(obsdat[j,2]))
  {track.na = cbind(track.na,j)}
}
track.na = track.na[-1]
if(length(track.na > 0)){
  obsdat = obsdat[-track.na,]
  size = size-length(track.na)}
r_x <-rep(0,size)
w_x <-rep(0,size)

for (j in 1:size)
{
  if(j==size){r_x[j]<-NA; w_x[j]<-NA }
  else{r_x[j]<-obsdat[j+1,1]*obsdat[j+1,2]/obsdat[j,2]
```

Ref. code: 25595809030157UGW

```
  w_x[j]<-((r_x[j])^2*((1/obsdat[j+1,2])+(1/obsdat[j,2])))^(-1)}
} ## end loop j

ex_na <- na.omit(r_x)
pnt_no <- length(ex_na)
if(pnt_no > 2){

  mod4 <- rq(log(r_x)~(obsdat[,1]+1),weights=w_x,na.action = na.omit)
  sum <- summary.rq(mod4,se="boot")
  alpha_hat4<- sum$coefficients[1]
  beta_hat4 <- sum$coefficients[2]
  varalpha4=sum$coefficients[3]^2
  varbeta4=sum$coefficients[4]^2
  if((alpha_hat4>-1)&&(beta_hat4>-Inf)&&(beta_hat4<Inf))
  {

    #newLB#
    A[i] <- alpha_hat4+beta_hat4
    f0 <- f1*exp(-A[i])
    Nhat.newLB[i] <- n+f0
    NLB1=(n*f0)/(n+f0)
    B1=exp(-A[i])^2
    B2=1-(f1/Nhat.newLB[i])
    B3=(f1)*B2
    NLB2=B3*B1
    B4=(f1^2)*B1
    NLB3=B4*(varalpha4+varbeta4)
    VarNhat.newLB[i]=NLB1+NLB2+NLB3

    i <- i+1
  }
```

```
  }
  else{ i <- i}
}##end B replicates

mean_A <- mean(A)
meannewLB <- mean(Nhat.newLB)
Rbias.newLB=(meannewLB-N)/N
RVar.newLB=mean((Nhat.newLB-meannewLB)^2,na.rm=TRUE)/N^2
RMSE.newLB=mean((Nhat.newLB-N)^2,na.rm=TRUE)/N^2
varNhatnewLB=mean(VarNhat.newLB)
seNhatnewLB=sqrt(varNhatnewLB)
 outcome<-cbind(meannewLB,Rbias.newLB,RVar.newLB,RMSE.newLB,
,varNhatnewLB,seNhatnewLB)
return(outcome)
} ##end program

cal1<-function(mu)
{library(VGAM)
 sink("d:\\result.xls")

cat("mu",'\t',"N",'\t',"k",'\t',"p",'\t','\t',"Truevar",'\t',"GTvar",'\t',"Cvar",'\t',"Zvar",'\t',"LBvar",'\n')
N<-c(5000,10000,50000,100000)
k<-c(5,7)
p<-rep(0,length(k))
for(i in 1:length(N)){
  for(j in 1:length(k)){
    p[j]<-k[j]/(k[j]+mu)
    result<-estimates(10000,N[i],k[j],p[j])

    cat(mu,'\t',N[i],'\t',k[j],'\t',p[j],'\t',alpha,'\t',result)
```

```
        }#end j
      }#end i
    }#endprogram
##################################################################
   ## For generate random number of Contaminated Poisson distribution ##
##################################################################

i<- 1
 set.seed(12345)
 while(i<=B)
 {
  rmixpois<- function(N){
   theta <- c(0.5,0.5)
   Y <- sample(2, N, replace = TRUE, prob = theta)
   x <- numeric(length = length(Y))
   for(i in 1:length(Y))
   {
    if(Y[i]==1){x[i] <- rpois(para1,para2)}
    if(Y[i]==2){x[i] <- rpois(para1,para2)}
   }
   return(x)
  }

  if (dist=="pois") {x <- rpois(N,para)}
  else if (dist=="geom") {x <- rgeom(N,para)}
  else if (dist=="nbinom") {x <- rnbinom(N,para[1],para[2])}
  else if (dist=="rmixpois"){x=rmixpois(N)}

##################################################################
   ## For generate random number of Contaminated Geometric distribution ##
##################################################################
```

Ref. code: 25595809030157UGW

```
i<- 1
  set.seed(12345)
  while(i<=B)
  {
    rmixgeo<- function(N){
      theta <- c(0.5,0.5)
      Y <- sample(2, N, replace = TRUE, prob = theta)
      x <- numeric(length = length(Y))
      for(i in 1:length(Y))
      {
        if(Y[i]==1){x[i] <- rgeom(para1,para2)}
        if(Y[i]==2){x[i] <- rgeom(para1,para2)}
      }
      return(x)
    }

    if (dist=="pois") {x <- rpois(N,para)}
    else if (dist=="geom") {x <- rgeom(N,para)}
    else if (dist=="nbinom") {x <- rnbinom(N,para[1],para[2])}
    else if (dist=="rmixgeo"){x=rmixgeo(N)}
```

2. เปรียบเทียบประสิทธิภาพของช่วงเชื่อมั่นของขนาดประชากร

การศึกษาใช้การจำลองข้อมูลและประมวลผล โดยการเขียนคำสั่งในโปรแกรม R Studio โดยมีคำสั่ง แสดงได้ดังนี้

```
###################################################################
## For generate random number of Poisson distribution, Geometric distribution and
Negative Binomial distribution ##
###################################################################

library(quantreg)
getdata<-function(getx,getfx)   ## adjust for freq = 0
{
  xlist <- c(0:max(getx))
  if(length(getx)==length(xlist)) { outx <-
matrix(cbind(getx,getfx),nrow=length(getx),ncol=2,byrow=FALSE)}
  else {
    "%w/o%" <- function(u,v) u[!u %in% v] #--  u without v
    xmis<- xlist %w/o% getx
    fmis <- rep(NA,length(xmis))
    mis_fig<-cbind(xmis,fmis)
    obs_fig<-cbind(getx,getfx)
    com_list<-rbind(obs_fig,mis_fig)
    aa<-com_list[,1]
    bb<-com_list[,2]
    o<-order(aa)
    outx <- matrix(cbind(aa[o],bb[o]),nrow=length(xlist),ncol=2,byrow=FALSE)}
  return(outx)
}

### main program ##
est.GTsimple<-function(B,N,dist,para)
```

```
{
  khat<-rep(0,B)
  Nhat.GT <-rep(0,B)
  VarNhat.GT <-rep(0,B)
  lower.NhatGT=rep(0,B)
  upper.NhatGT=rep(0,B)
  length.NhatGT=rep(0,B)
  tempGT=rep(0,B)

  i<- 1
  set.seed(12345)
  while(i<=B)
  {
    if (dist=="pois") {x <- rpois(N,para)}
    else if (dist=="geom") {x <- rgeom(N,para)}
    else if (dist=="nbinom") {x <- rnbinom(N,para[1],para[2])}
    s <- sum(x)
    f1 <-sum(as.numeric(x==1))
    f2 <- sum(as.numeric(x==2))
    f3 <- sum(as.numeric(x==3))
    tab<-table(x)
    dtab <- as.data.frame(tab)
    out<-attributes(tab)
    outx<-out$dimnames$x
    getx<-as.numeric(outx)
    getfx<- as.numeric(dtab$Freq)  ## get fx
    size<-max(getx)
    xdat <- getdata(getx,getfx)
    obsdat <- xdat[-1,]  ### observed frequency
    n <- sum(obsdat[,2],na.rm=TRUE)
    newsize = size
```

Ref. code: 25595809030157UGW

```
track.na = 0
for(j in 1:size){
  if(is.na(obsdat[j,2]))
  {track.na = cbind(track.na,j)}
}
track.na = track.na[-1]
if(length(track.na > 0)){
  obsdat = obsdat[-track.na,]
  size = size-length(track.na)}
r_x <-rep(0,size)
w_x <-rep(0,size)

for (j in 1:size)
{
  if(j==size){r_x[j]<-NA; w_x[j]<-NA }
  else{r_x[j]<-obsdat[j+1,1]*obsdat[j+1,2]/obsdat[j,2]
  w_x[j]<-((r_x[j])^2*((1/obsdat[j+1,2])+(1/obsdat[j,2])))^(-1)}
} ## end loop j

ex_na <- na.omit(r_x)
pnt_no <- length(ex_na)
if(pnt_no > 2){
  mod1 <- lm(r_x~obsdat[,1],weights=w_x)
  sum <- summary(mod1)
  alpha_hat1 <- sum$coefficients[1]
  beta_hat1 <- sum$coefficients[2]

  if((alpha_hat1>0)&&(beta_hat1>0)&&(beta_hat1<1))
  {
    #GT#
    khat[i] <- alpha_hat1/beta_hat1
```

Ref. code: 25595809030157UGW

```
        p0 <- (f1/s)^(khat[i]/(khat[i]+1))
        Nhat.GT[i] <- n/(1-p0)
        Var1<-(n*p0)/((1-p0)^2)
        A1 <- (n^2)/((1-p0)^4)
        A2 <- s^((-2*khat[i])/(khat[i]+1))
        A3 <- (khat[i]/(khat[i]+1))^2
        A4 <- f1^((khat[i]-1)/(khat[i]+1))
        A5 <- 1-(f1/Nhat.GT[i])
        A6 <- s^((-3*khat[i])/(khat[i]+1))
        A7 <- f1^(((2*khat[i])-1)/(khat[i]+1))
        Var2<-  A1*((A2*A3*A4*A5)+(A6*A3*A7))
        VarNhat.GT[i] <- Var1+Var2
        lower.NhatGT[i]=Nhat.GT[i]-qnorm(1-0.025)*sqrt(VarNhat.GT[i] )
        upper.NhatGT[i]=Nhat.GT[i]+qnorm(1-0.025)*sqrt(VarNhat.GT[i] )
        if((N>=lower.NhatGT[i])&(N<=upper.NhatGT[i])){
tempGT[i]=1
        length.NhatGT[i]=upper.NhatGT[i]-lower.NhatGT[i]}
        else{tempGT[i]=0}
        i <- i+1}
    }
    else{ i <- i}
  }##end B replicates
  mean_khat <- mean(khat)
  meanNGT <- mean(Nhat.GT)

  conf.NhatGT=mean(tempGT)
  lengthNhatGT=mean(length.NhatGT)
  VarNhatGT=mean(VarNhat.GT)
  seNhatGT=sqrt(VarNhatGT)

  outcome<-cbind(meanNGT,conf.NhatGT,lengthNhatGT,VarNhatGT,seNhatGT)
```

```
  return(outcome)
} ##end program

est.GTquantile<-function(B,N,dist,para)
{
  khat<-rep(0,B)
  Nhat.newGT <-rep(0,B)
  VarNhat.newGT <-rep(0,B)
  lower.NhatnewGT=rep(0,B)
  upper.NhatnewGT=rep(0,B)
  length.NhatnewGT=rep(0,B)
  tempnewGT=rep(0,B)

  i<- 1
  set.seed(12345)
  while(i<=B)
  {
    if (dist=="pois") {x <- rpois(N,para)}
    else if (dist=="geom") {x <- rgeom(N,para)}
    else if (dist=="nbinom") {x <- rnbinom(N,para[1],para[2])}
    s <- sum(x)
    f1 <-sum(as.numeric(x==1))
    f2 <- sum(as.numeric(x==2))
    f3 <- sum(as.numeric(x==3))
    tab<-table(x)
    dtab <- as.data.frame(tab)
    out<-attributes(tab)
    outx<-out$dimnames$x
    getx<-as.numeric(outx)
    getfx<- as.numeric(dtab$Freq)  ## get fx
    size<-max(getx)
```

```
xdat <- getdata(getx,getfx)
obsdat <- xdat[-1,]  ### observed frequency
n <- sum(obsdat[,2],na.rm=TRUE)
newsize = size
track.na = 0
for(j in 1:size){
  if(is.na(obsdat[j,2]))
  {track.na = cbind(track.na,j)}
}
track.na = track.na[-1]
if(length(track.na > 0)){
  obsdat = obsdat[-track.na,]
  size = size-length(track.na)}
r_x <-rep(0,size)
w_x <-rep(0,size)

for (j in 1:size)
{
  if(j==size){r_x[j]<-NA; w_x[j]<-NA }
  else{r_x[j]<-obsdat[j+1,1]*obsdat[j+1,2]/obsdat[j,2]
  w_x[j]<-((r_x[j])^2*((1/obsdat[j+1,2])+(1/obsdat[j,2])))^(-1)}
} ## end loop j

ex_na <- na.omit(r_x)
pnt_no <- length(ex_na)
if(pnt_no > 2){ mod3 <- rq(r_x~obsdat[,1],weights=w_x,na.action = na.omit)
sum <- summary.rq(mod3)
alpha_hat3<- sum$coefficients[1]
beta_hat3 <- sum$coefficients[2]
if((alpha_hat3>0)&&(beta_hat3>0)&&(beta_hat3<1))
```

```
    {
      #newGT#
      khat[i] <- alpha_hat3/beta_hat3
      p0 <- (f1/s)^(khat[i]/(khat[i]+1))
      Nhat.newGT[i] <- n/(1-p0)
      Var1<-(n*p0)/((1-p0)^2)
      A1 <- (n^2)/((1-p0)^4)
      A2 <- s^((-2*khat[i])/(khat[i]+1))
      A3 <- (khat[i]/(khat[i]+1))^2
      A4 <- f1^((khat[i]-1)/(khat[i]+1))
      A5 <- 1-(f1/Nhat.newGT[i])
      A6 <- s^((-3*khat[i])/(khat[i]+1))
      A7 <- f1^(((2*khat[i])-1)/(khat[i]+1))
      Var2<-  A1*((A2*A3*A4*A5)+(A6*A3*A7))
      VarNhat.newGT[i] <- Var1+Var2
      lower.NhatnewGT[i]=Nhat.newGT[i]-qnorm(1-0.025)*sqrt(VarNhat.newGT[i] )
      upper.NhatnewGT[i]=Nhat.newGT[i]+qnorm(1-0.025)*sqrt(VarNhat.newGT[i] )
      if((N>=lower.NhatnewGT[i])&(N<=upper.NhatnewGT[i])){
tempnewGT[i]=1
      length.NhatnewGT[i]=upper.NhatnewGT[i]-lower.NhatnewGT[i]
}
      else{tempnewGT[i]=0}
      i <- i+1}
    }
    else{ i <- i}
  }##end B replicates

  mean_khat <- mean(khat)
  meannewGT <- mean(Nhat.newGT)
  conf.NhatnewGT=mean(tempnewGT)
  lengthNhatnewGT=mean(length.NhatnewGT)
```

```
  VarNhatnewGT=mean(VarNhat.newGT)
  seNhatnewGT=sqrt(VarNhatnewGT)
outcome<-
cbind(meannewGT,conf.NhatnewGT,lengthNhatnewGT,VarNhatnewGT,seNhatnewGT)
  return(outcome)
} ##end program

est.LBsimple<-function(B,N,dist,para)
{
  Nhat.LB <-rep(0,B)
  VarNhat.LB <-rep(0,B)
  lower.NhatLB=rep(0,B)
  upper.NhatLB=rep(0,B)
  length.NhatLB=rep(0,B)
  tempLB=rep(0,B)

  i<- 1
  set.seed(12345)
  while(i<=B)
  {
    if (dist=="pois") {x <- rpois(N,para)}
    else if (dist=="geom") {x <- rgeom(N,para)}
    else if (dist=="nbinom") {x <- rnbinom(N,para[1],para[2])}
    s <- sum(x)
    f1 <-sum(as.numeric(x==1))
    f2 <- sum(as.numeric(x==2))
    f3 <- sum(as.numeric(x==3))
    tab<-table(x)
    dtab <- as.data.frame(tab)
    out<-attributes(tab)
    outx<-out$dimnames$x
```

```
getx<-as.numeric(outx)
getfx<- as.numeric(dtab$Freq)  ## get fx
size<-max(getx)
xdat <- getdata(getx,getfx)
obsdat <- xdat[-1,]  ### observed frequency
n <- sum(obsdat[,2],na.rm=TRUE)
newsize = size
track.na = 0
for(j in 1:size){
  if(is.na(obsdat[j,2]))
  {track.na = cbind(track.na,j)}
}
track.na = track.na[-1]
if(length(track.na > 0)){
  obsdat = obsdat[-track.na,]
  size = size-length(track.na)}
r_x <-rep(0,size)
w_x <-rep(0,size)
for (j in 1:size)
{
  if(j==size){r_x[j]<-NA; w_x[j]<-NA }
  else{r_x[j]<-obsdat[j+1,1]*obsdat[j+1,2]/obsdat[j,2]
  w_x[j]<-((r_x[j])^2*((1/obsdat[j+1,2])+(1/obsdat[j,2])))^(-1)}
} ## end loop j

ex_na <- na.omit(r_x)
pnt_no <- length(ex_na)
if(pnt_no > 2){ mod1 <- lm(log(r_x)~(obsdat[,1]+1),weights=w_x)
sum <- summary(mod1)
alpha_hat1<- sum$coefficients[1]
beta_hat1 <- sum$coefficients[2]
```

```
    if((alpha_hat1>-1)&&(beta_hat1>-Inf)&&(beta_hat1<Inf))
    {
      #LB#
      Nhat.LB[i]<- n+(3*f3*f1^3)/(4*f2^3)
      LB1<-(9/4)^2*(f1^5*f3^2/f2^6)*(f1/f2+1)
      LB2<-(3/4)^2*(f1^6*f3/f2^6)*(1-f3/n)
      LB3<-(3*n/4)*(f1^3*f3)/(n*f2^3+3*f1^3*f3/4)
      VarNhat.LB[i]<-LB1+LB2+LB3
      lower.NhatLB[i]=Nhat.LB[i]-qnorm(1-0.025)*sqrt(VarNhat.LB[i] )
      upper.NhatLB[i]=Nhat.LB[i]+qnorm(1-0.025)*sqrt(VarNhat.LB[i] )
      if((N>=lower.NhatLB[i])&(N<=upper.NhatLB[i])){
tempLB[i]=1
      length.NhatLB[i]=upper.NhatLB[i]-lower.NhatLB[i]
}
      else{tempLB[i]=0}
      i <- i+1}
    }
    else{ i <- i}
  }##end B replicates
  meanNLB <- mean(Nhat.LB)
  conf.NhatLB=mean(tempLB)
  lengthNhatLB=mean(length.NhatLB)
  varNhatLB=mean(VarNhat.LB)
  seNhatLB=sqrt(varNhatLB)

  outcome<-cbind(meanNLB,conf.NhatLB,lengthNhatLB,varNhatLB,seNhatLB)
  return(outcome)
} ##end program
est.LBquantile<-function(B,N,dist,para)
{
  A<-rep(0,B)
```

```
Nhat.newLB <-rep(0,B)
VarNhat.newLB <-rep(0,B)
lower.NhatnewLB=rep(0,B)
upper.NhatnewLB=rep(0,B)
length.NhatnewLB=rep(0,B)
tempnewLB=rep(0,B)

i<- 1
set.seed(12345)
while(i<=B)
{
  if (dist=="pois") {x <- rpois(N,para)}
  else if (dist=="geom") {x <- rgeom(N,para)}
  else if (dist=="nbinom") {x <- rnbinom(N,para[1],para[2])}
  s <- sum(x)
  f1 <-sum(as.numeric(x==1))
  f2 <- sum(as.numeric(x==2))
  f3 <- sum(as.numeric(x==3))
  tab<-table(x)
  dtab <- as.data.frame(tab)
  out<-attributes(tab)
  outx<-out$dimnames$x
  getx<-as.numeric(outx)
  getfx<- as.numeric(dtab$Freq)  ## get fx
  size<-max(getx)
  xdat <- getdata(getx,getfx)
  obsdat <- xdat[-1,]  ### observed frequency
  n <- sum(obsdat[,2],na.rm=TRUE)
  newsize = size
  track.na = 0
  for(j in 1:size){
```

```
  if(is.na(obsdat[j,2]))
  {track.na = cbind(track.na,j)}
}
track.na = track.na[-1]
if(length(track.na > 0)){
  obsdat = obsdat[-track.na,]
  size = size-length(track.na)}
r_x <-rep(0,size)
w_x <-rep(0,size)

for (j in 1:size)
{
  if(j==size){r_x[j]<-NA; w_x[j]<-NA }
  else{r_x[j]<-obsdat[j+1,1]*obsdat[j+1,2]/obsdat[j,2]
  w_x[j]<-((r_x[j])^2*((1/obsdat[j+1,2])+(1/obsdat[j,2])))^(-1)}
} ## end loop j

ex_na <- na.omit(r_x)
pnt_no <- length(ex_na)
if(pnt_no > 2){

  mod4 <- rq(log(r_x)~(obsdat[,1]+1),weights=w_x,na.action = na.omit)
  sum <- summary.rq(mod4,se="boot")
  alpha_hat4<- sum$coefficients[1]
  beta_hat4 <- sum$coefficients[2]
  varalpha4=sum$coefficients[3]^2
  varbeta4=sum$coefficients[4]^2
  if((alpha_hat4>-1)&&(beta_hat4>-Inf)&&(beta_hat4<Inf))
  {

    #newLB#
```

Ref. code: 25595809030157UGW

```
        A[i] <- alpha_hat4+beta_hat4
        f0 <- f1*exp(-A[i])
        Nhat.newLB[i] <- n+f0
        NLB1=(n*f0)/(n+f0)
        B1=exp(-A[i])^2
        B2=1-(f1/Nhat.newLB[i])
        B3=(f1)*B2
        NLB2=B3*B1
        B4=(f1^2)*B1
        NLB3=B4*(varalpha4+varbeta4)
        VarNhat.newLB[i]=NLB1+NLB2+NLB3
        lower.NhatnewLB[i]=Nhat.newLB[i]-qnorm(1-0.025)*sqrt(VarNhat.newLB[i])
        upper.NhatnewLB[i]=Nhat.newLB[i]+qnorm(1-0.025)*sqrt(VarNhat.newLB[i] )
        if((N>=lower.NhatnewLB[i])&&(N<=upper.NhatnewLB[i])){
tempnewLB[i]=1
        length.NhatnewLB[i]=upper.NhatnewLB[i]-lower.NhatnewLB[i]
}
        else{tempnewLB[i]=0}
        i <- i+1
      }
    }
    else{ i <- i}
  }##end B replicates

  mean_A <- mean(A)
  meannewLB <- mean(Nhat.newLB)
  conf.NhatnewLB=mean(tempnewLB)
  lengthNhatnewLB=mean(length.NhatnewLB)
  varNhatnewLB=mean(VarNhat.newLB)
  seNhatnewLB=sqrt(varNhatnewLB)
```

```
  outcome<-
cbind(meannewLB,conf.NhatnewLB,lengthNhatnewLB,varNhatnewLB,seNhatnewLB)
  return(outcome)
} ##end program

cal1<-function(mu)
{library(VGAM)
  sink("d:\\result.xls")

cat("mu",'\t',"N",'\t',"k",'\t',"p",'\t','\t',"Truevar",'\t',"GTvar",'\t',"Cvar",'\t',"Zvar",'\t',"LBvar",'\n')
  N<-c(5000,10000,50000,100000)
  k<-c(5,7)
  p<-rep(0,length(k))
  for(i in 1:length(N)){
    for(j in 1:length(k)){
      p[j]<-k[j]/(k[j]+mu)
      result<-estimates(10000,N[i],k[j],p[j])

      cat(mu,'\t',N[i],'\t',k[j],'\t',p[j],'\t',alpha,'\t',result)

    }#end j
  }#end i
}#endprogram

###################################################################
      ## For generate random number of Contaminated Poisson distribution ##
###################################################################

i<- 1
  set.seed(12345)
  while(i<=B)
```

```
 {
   rmixpois<- function(N){
     theta <- c(0.5,0.5)
     Y <- sample(2, N, replace = TRUE, prob = theta)
     x <- numeric(length = length(Y))
     for(i in 1:length(Y))
     {
       if(Y[i]==1){x[i] <- rpois(para1,para2)}
       if(Y[i]==2){x[i] <- rpois(para1,para2)}
     }
     return(x)
   }

   if (dist=="pois") {x <- rpois(N,para)}
   else if (dist=="geom") {x <- rgeom(N,para)}
   else if (dist=="nbinom") {x <- rnbinom(N,para[1],para[2])}
   else if (dist=="rmixpois"){x=rmixpois(N)}

####################################################################
      ## For generate random number of Contaminated Geometric distribution ##
####################################################################

i<- 1
 set.seed(12345)
 while(i<=B)
 {
   rmixgeo<- function(N){
     theta <- c(0.5,0.5)
     Y <- sample(2, N, replace = TRUE, prob = theta)
     x <- numeric(length = length(Y))
     for(i in 1:length(Y))
```

Ref. code: 25595809030157UGW

```
  {
    if(Y[i]==1){x[i] <- rgeom(para1,para2)}
    if(Y[i]==2){x[i] <- rgeom(para1,para2)}
  }
  return(x)
}

if (dist=="pois") {x <- rpois(N,para)}
else if (dist=="geom") {x <- rgeom(N,para)}
else if (dist=="nbinom") {x <- rnbinom(N,para[1],para[2])}
else if (dist=="rmixgeo"){x=rmixgeo(N)}
```

ประวัติผู้เขียน

ชื่อ	นางสาวลภัสรดา สิงห์สมบูรณ์
วันเดือนปีเกิด	21 พฤษภาคม 2534
การศึกษา	ปีการศึกษา 2557: วิทยาศาสตรบัณฑิต (สถิติ) มหาวิทยาลัยบูรพา
ประสบการณ์ทำงาน	2557-2558 เจ้าหน้าที่ดูแลลูกค้า ธนาคารกรุงไทย สาขาบิ๊กซีชลบุรี

Ref. code: 25595809030157UGW