

การประมาณจำนวนผู้ติดเชื้อโควิด-19 ในกรุงเทพมหานคร โดยการใช้วิธีจับ-จับใหม่

Estimating the Number of COVID-19 Infected Cases in Bangkok by Capture-Recapture Method

ปารวัน ไพจิตรรัตน์* กรชนก ป้องกัน, ศศิชา จารุนทร และ ชญาดา ยิ่งเจริญธนา

Parawan Pijitrattana* Kornchanok Pongkan, Sasicha Jarunet and Chayada Yingcharoenthana

สาขาวิชาคณิตศาสตร์และสถิติ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์ ประเทศไทย

Department of Mathematics and Statistics, Faculty of Science and Technology, Thammasat University, Thailand

บทคัดย่อ

วิธีการจับ-จับใหม่ ใช้ในการประมาณขนาดประชากรที่ยากต่อการสำรวจพบ ซึ่งปัจจุบันมีการนำไปใช้อย่างแพร่หลายในงานด้านต่าง ๆ รวมถึงด้านระบาดวิทยา เนื่องจากผู้ติดเชื้อโควิด-19 ส่วนใหญ่มักจะไม่แสดงอาการ ทำให้มีผู้ติดเชื้อซ่อนเร้นอยู่เป็นจำนวนมาก งานวิจัยนี้ประยุกต์ใช้วิธีจับ-จับใหม่เพื่อประมาณจำนวนประชากรผู้ติดเชื้อโควิด-19 ในกรุงเทพมหานครเดือนพฤศจิกายน พ.ศ. 2564 ถึง มกราคม พ.ศ. 2565 โดยใช้ตัวประมาณขนาดประชากรจากการแจกแจงเรขาคณิตซึ่งเป็นตัวประมาณขอบเขตล่าง และตัวประมาณขอบเขตบนภายใต้การแจกแจงเรขาคณิต เมื่อพิจารณาตัวประมาณขอบเขตล่างพบว่าในเดือนพฤศจิกายน ธันวาคม และมกราคม มีผู้ติดเชื้อจำนวน 46,134, 35,427 และ 58,083 ราย ตามลำดับ คิดเป็น 2 เท่าของจำนวนผู้ติดเชื้อที่มีการรายงานโดยกรมควบคุมโรค สำหรับตัวประมาณขอบเขตบนพบว่าในเดือนพฤศจิกายน ธันวาคม และมกราคม มีจำนวนผู้ติดเชื้อจำนวน 99,258, 76,443 และ 125,377 ราย ตามลำดับ คิดเป็น 4.5 เท่าของจำนวนผู้ติดเชื้อที่มีการรายงานโดยกรมควบคุมโรค

คำสำคัญ: การประมาณขนาดประชากร ไวรัสโคโรนา ไร้อาการ ตรวจไม่พบ

ABSTRACT

Capture-recapture method is widely used in a variety of fields including epidemiology to estimate the size of difficult-to-explore populations. Many people are asymptomatic after contracting COVID-19, resulting in a significant number of concealed infections. This study applied capture-recapture method to estimate the total number of infected populations in Bangkok from November 2021 to January 2022 using Chao lower bound estimator and upper bound estimator under geometric distribution. When the lower bound estimates were taken into account, it was discovered that there were 46,134, 35,427, and 58,083 cases in November, December, and January, respectively, representing 2 times the number of infected reported by the Department of Disease Control. According to the upper bound estimator, there were 99,258, 76,443, and 125,377 cases in November, December, and January, respectively, representing 4.5 times the number of infections reported by the Department of Disease Control.

KEYWORDS: population size estimation, coronavirus, asymptomatic, undetected

*Corresponding Author: parawan@mathstat.sci.tu.ac.th

Received: 18/04/2022; Revised: 10/02/2023; Accepted: 14/02/2023

1. บทนำ

การระบาดของเชื้อไวรัสโคโรนาหรือโควิด-19 เริ่มระบาดในเมืองอุฮั่น มณฑลหูเป่ย์ของประเทศจีน ในช่วงปลายปี พ.ศ. 2562 ได้มีการระบาดไปยังประเทศต่าง ๆ ทั่วโลก ไวรัสโคโรนา (coronavirus) เป็นไวรัสที่ถูกพบครั้งแรกในปี ค.ศ. 1960 แต่ยังไม่ทราบแหล่งที่มาอย่างชัดเจน ไวรัสนี้สามารถติดเชื้อได้ทั้งในมนุษย์และสัตว์

ปัจจุบันมีการค้นพบไวรัสสายพันธุ์นี้แล้วทั้งหมด 6 สายพันธุ์ ส่วนสายพันธุ์ที่กำลังแพร่ระบาดในปัจจุบันถูกเรียกว่า “ไวรัสโคโรนาสายพันธุ์ใหม่” มีชื่อทางการว่า “โควิด-19 (coronavirus disease starting in 2019: COVID-19)” โดยเชื้อโควิด-19 สามารถแพร่ผ่านทางละอองน้ำมูก น้ำลาย ผ่านการไอ จาม และสามารถอยู่บนพื้นผิวของวัตถุต่าง ๆ ทำให้ผู้ที่สัมผัสอาจรับเชื้อเข้าสู่

ร่างกายได้ โดยผู้ติดเชื้อโควิด-19 อาจมีอาการเริ่มต้น เช่น ไอแห้ง หายใจไม่สะดวก ไม่สามารถรับรู้รสชาติและกลิ่นได้ และมีผู้ติดเชื้อจำนวนมากไม่แสดงอาการ

การแพร่ระบาดของโควิด-19 ในประเทศไทยมีความรุนแรงมากขึ้นนับตั้งแต่เริ่มมีการระบาด มีจำนวนผู้ป่วยและจำนวนผู้เสียชีวิตเพิ่มมากขึ้น แต่ข้อมูลจำนวนผู้ติดเชื้อโควิด-19 จากการรายงานของภาครัฐอาจไม่ใช่ข้อมูลที่ถูกต้อง เนื่องจากมีผู้ติดเชื้อที่มีอาการไม่รุนแรงหรือไม่แสดงอาการอยู่เป็นจำนวนมาก ทำให้มีผู้ติดเชื้อซ่อนเร้นที่ไม่ถูกบันทึกเข้าสู่ระบบสาธารณสุข ซึ่งอาจก่อให้เกิดปัญหาในการเฝ้าระวังทางสาธารณสุขในเรื่องของการเตรียมทรัพยากรทางสาธารณสุขเพื่อรองรับผู้ป่วย และการวางแผนในการป้องกันการติดเชื้อและดำเนินการเกี่ยวกับผู้ที่ได้รับเชื้อ

วิธีจับ-จับใหม่ (capture-recapture method) เป็นวิธีที่ใช้ประมาณขนาดประชากรที่หน่วยตัวอย่างยากต่อการพบเจอ แต่เดิมวิธีจับ-จับใหม่ ใช้ในการประมาณขนาดประชากรสัตว์ป่าที่อาศัยอยู่ในพื้นที่กว้าง มีการหลบซ่อนและมีการเคลื่อนย้ายถิ่นฐานตลอดเวลา โดยทำการสุ่มตัวอย่างครั้งแรกจากประชากรและทำเครื่องหมายแล้วปล่อยกลับคืน หลังจากนั้นรอรยะเวลาหนึ่งจึงทำการสุ่มตัวอย่างจากประชากรเดิมอีกครั้งและทำเครื่องหมายซ้ำ ๆ กันหลายครั้ง (repeated identification) ซึ่งหน่วยตัวอย่างที่ได้จากการสุ่มในแต่ละครั้งอาจมีหน่วยตัวอย่างซ้ำจากการสุ่มครั้งก่อน ๆ ที่ถูกทำเครื่องหมายไว้ ปัจจุบันวิธีจับ-จับใหม่ถูกประยุกต์ใช้อย่างแพร่หลาย ทั้งในด้านชีววิทยา สังคมวิทยาและอาชีววิทยา รวมไปถึงทางด้านสาธารณสุขและด้านระบาดวิทยา เช่น การประมาณจำนวนผู้เสพยาเสพติดของจังหวัดเชียงใหม่ในปี พ.ศ 2559-2560 (Pijitratana *et.al.*, 2020) การประมาณจำนวนผู้ติดเชื้อเอชไอวีในประเทศไทย (Lanumtaeng, &Bohning, 2011) การประมาณจำนวนโสเภณีในประเทศเวียดนาม (Nguyen *et.al.*, 2021) การตรวจสอบความสมบูรณ์ของการสอบสวนโรคและการติดตามผู้สัมผัสโรคโควิด-19 ในช่วงการระบาดระลอกแรกของประเทศไทย (Lerdsuwansri *et.al.*, 2022)

ข้อมูลที่ได้จากการเก็บรวบรวมด้วยวิธีจับ-จับใหม่เป็นข้อมูลนับ (count data) โดยสนใจจำนวนครั้งที่หน่วยตัวอย่างถูกพบในช่วงเวลาที่ทำการศึกษา ถ้ากำหนดให้ตัวแปรสุ่ม Y แทน จำนวนครั้งที่หน่วยตัวอย่างถูกพบในช่วงเวลาที่ทำการศึกษา โดย $y=0, 1, 2, 3, \dots, m$ และ f_i แทน จำนวนหน่วย

ตัวอย่างที่ถูกพบ i ครั้ง โดยที่ $i = 0, 1, 2, \dots, m$ เมื่อ m คือ จำนวนครั้งสูงสุดที่พบ เนื่องจาก f_0 คือ จำนวนหน่วยตัวอย่างที่ไม่ถูกพบ จึงไม่ทราบค่า f_0 ถ้ากำหนดให้ความน่าจะเป็นที่แต่ละหน่วยตัวอย่างในประชากรจะไม่ถูกพบเท่ากับ p_0 ดังนั้นความน่าจะเป็นที่แต่ละหน่วยตัวอย่างในประชากรจะถูกพบเท่ากับ $1-p_0$ และ N แทน ขนาดประชากรที่ไม่ทราบค่าซึ่งประกอบด้วยจำนวนหน่วยตัวอย่างที่ถูกพบและไม่ถูกพบ

$$N = n + f_0 = n + Np_0 \quad (1)$$

เมื่อ n แทนจำนวนหน่วยตัวอย่างที่ถูกพบโดย $n = f_1 + f_2 + \dots + f_m$ จากสมการ (1) จะได้ตัวประมาณขนาดประชากร Horvitz-Thompson (Van Der Heijden *et.al.*, 2003)

$$\hat{N} = \frac{n}{1 - \hat{p}_0} \quad (2)$$

ดังนั้นในการประมาณขนาดประชากรต้องทำการประมาณค่า p_0 ซึ่งโดยส่วนใหญ่มักอธิบายข้อมูลนับด้วยการแจกแจงปัวซอง (Poisson distribution) ที่มีพารามิเตอร์ λ โดยความน่าจะเป็นที่หน่วยตัวอย่างแต่ละหน่วยจะถูกพบ i ครั้งมีค่าเท่ากัน คือ $\frac{e^{-\lambda} \lambda^i}{i!}$ เรียกว่าตัวแบบปัวซองเอกพันธ์ (homogeneous Poisson model) แต่ในความเป็นจริง หน่วยตัวอย่างในประชากรมักมีความหลากหลาย โดยประชากรประกอบด้วยกลุ่มประชากรย่อย (subpopulation) ซึ่งประชากรย่อยแต่ละกลุ่มมีการแจกแจงปัวซองที่มีพารามิเตอร์ Λ มีค่าแตกต่างกัน ทำให้ความน่าจะเป็นที่หน่วยตัวอย่างแต่ละหน่วยจะถูกพบ ครั้งมีค่าไม่เท่ากัน กรณีเช่นนี้เรียกว่าตัวแบบปัวซองวิวิธพันธุ์ (heterogeneous Poisson model) โดยฟังก์ชันการแจกแจงของตัวแบบปัวซองวิวิธพันธุ์ คือ $p_i = \int_0^\infty \frac{e^{-\lambda} \lambda^i}{i!} f(\lambda) d\lambda$

เมื่อพารามิเตอร์ Λ ของการแจกแจงปัวซองมีการแจกแจงแบบเลขชี้กำลัง (exponential distribution) เกิดเป็นการแจกแจงผสมปัวซองเลขชี้กำลังหรือการแจกแจงเรขาคณิต (geometric distribution)

ในปี พ.ศ. 2563 Böhning และคณะ (Böhning *et.al.*, 2020) ได้ประมาณจำนวนผู้ติดเชื้อโควิด-19 ในประเทศออสเตรเลีย จากข้อมูลจำนวนผู้ติดเชื้อสะสมและจำนวนผู้เสียชีวิตสะสม Böhning และคณะ สมมติให้จำนวนผู้ติดเชื้อโควิด-19 มีการแจกแจงเรขาคณิต และใช้ตัวประมาณ

ชาวภายใต้การแจกแจงเรขาคณิต ที่นำเสนอโดย Niwitpong และคณะ ในปี พ.ศ. 2556 (Niwitpong *et.al.*, 2013) มีสูตรในการคำนวณคือ $\hat{N}_{Chao} = n + \hat{f}_0 = n + \frac{f_1^2}{f_2}$ ในการ

ประมาณขนาดประชากรผู้ติดเชื้อโควิด-19 ซึ่งเป็นตัวประมาณขอบเขตล่าง อีกทั้ง Niwitpong และคณะยังได้เสนอตัวประมาณความแปรปรวนของตัวประมาณชาวภายใต้การแจกแจงเรขาคณิต มีสูตรในการคำนวณคือ $Var(\hat{N}_{Chao}) = \frac{f_1^4}{f_2^3} + \frac{4f_1^3}{f_2^2} + \frac{f_1^2}{f_2}$ โดย Böhning และคณะ

กำหนดให้ $f_1(t) = \Delta N(t)$ และ $f_2(t) = \Delta N(t-1) - \Delta D(t)$

$N(t)$ แทน จำนวนผู้ติดเชื้อสะสม ณ วันที่ t

$N(t-1)$ แทน จำนวนผู้ติดเชื้อสะสม ณ วันที่ $t-1$

$D(t)$ แทน จำนวนผู้เสียชีวิตสะสม ณ วันที่ $t-1$

$D(t-1)$ แทน จำนวนผู้เสียชีวิตสะสม ณ วันที่ t

$\Delta N(t) = N(t) - N(t-1)$ แทน จำนวนผู้ติดเชื้อ ณ วันที่ t

$\Delta D(t) = D(t) - D(t-1)$ แทน จำนวนผู้เสียชีวิต ณ วันที่ t

สามารถประมาณจำนวนผู้ติดเชื้อซ่อนเร้น ณ วันที่ t โดยตัวประมาณชาวภายใต้การแจกแจงเรขาคณิตจาก

$\hat{f}_0^C(t) = \frac{f_1^2(t)}{f_2(t)}$ เมื่อ $t = t_0 + 1, \dots, t_m$ และประมาณ

จำนวนผู้ติดเชื้อซ่อนเร้นในช่วงระยะเวลาที่ศึกษาโดย

$\hat{f}_0^C = \sum_{t=t_0+1}^{t_m} \hat{f}_0^C(t) = \sum_{t=t_0+1}^{t_m} \frac{f_1^2(t)}{f_2(t)}$ ดังนั้นจำนวนผู้ติดเชื้อ

ในช่วงที่ศึกษาคือ $\hat{N}_C = n + \sum_{t=t_0+1}^{t_m} \frac{f_1^2(t)}{f_2(t)}$ เมื่อ n แทน

จำนวนผู้ติดเชื้อที่มีการรายงานในช่วงที่ศึกษา อีกทั้ง Böhning และคณะได้ตรวจสอบประสิทธิภาพของตัวประมาณชาวภายใต้การแจกแจงเรขาคณิต โดยการคำนวณอัตราการติดเชื้อที่ประมาณได้จากตัวประมาณชาวภายใต้การแจกแจงเรขาคณิตเทียบกับอัตราการติดเชื้อจากการตรวจด้วยวิธี RT-PCR (real time PCR) พบว่ามีค่าใกล้เคียงกัน

ในปี พ.ศ. 2563 Rocchetti และคณะ (Rocchetti *et.al.*, 2020) ได้เสนอตัวประมาณขอบเขตบนภายใต้การแจกแจงเรขาคณิตเพื่อประมาณจำนวนผู้ติดเชื้อโควิด-19 ในทวีปยุโรป โดยประยุกต์ใช้ตัวประมาณขอบเขตบนของ Alfó และคณะ (Alfó *et.al.*, 2020) กับแนวคิดของ Böhning และคณะ (Böhning *et.al.*, 2020) ในการกำหนด $f_1(t) = \Delta N(t)$, $f_2(t) = \Delta N(t-1) - \Delta D(t)$ และ $f_3(t) = \Delta N(t-2) - \Delta D(t-1) - \Delta D(t)$ โดยประมาณจำนวนผู้ติดเชื้อซ่อนเร้น ณ วันที่ t จากสมการที่ (3)

$$\hat{f}_0^{UB}(t) = n^*(t) \left(\frac{\hat{\pi}_0^{UB}(t)}{1 - \hat{\pi}_0^{UB}(t)} \right), \quad (3)$$

$$n^*(t) = f_1(t) + f_2(t) + f_3(t),$$

$$\hat{\pi}_0^{UB}(t) = \frac{\frac{f_1^2(t)}{f_2(t)}}{f_1(t) + f_2(t) + \frac{f_1^2(t)}{f_2(t)}}$$

เมื่อ $t = t_0 + 2, \dots, t_m$ และประมาณจำนวนผู้ติดเชื้อซ่อนเร้นในช่วงระยะเวลาที่ทำการศึกษาโดย

$$\hat{f}_0^{UB} = \sum_{t=t_0+2}^{t_m} \hat{f}_0^{UB}(t) = \sum_{t=t_0+2}^{t_m} n^*(t) \left(\frac{\hat{\pi}_0^{UB}(t)}{1 - \hat{\pi}_0^{UB}(t)} \right)$$

ดังนั้นจำนวนผู้ติดเชื้อในช่วงที่ศึกษา คือ $\hat{N}_{UB} = n + \hat{f}_0^{UB}$ เมื่อ n แทนจำนวนผู้ติดเชื้อที่มีการรายงานในช่วงที่ศึกษา

งานวิจัยนี้ทำการประมาณขนาดประชากรผู้ติดเชื้อโควิด-19 ในกรุงเทพมหานครด้วยวิธีจับ-จับใหม่ โดยใช้ตัวประมาณขนาดประชากรที่พัฒนาภายใต้การแจกแจงปัวซองวิวิพันธ์ 2 ตัวประมาณ ได้แก่ ตัวประมาณชาวภายใต้การแจกแจงเรขาคณิตซึ่งเป็นตัวประมาณขอบเขตล่าง (Böhning *et.al.*, 2020) และตัวประมาณขอบเขตบนภายใต้การแจกแจงเรขาคณิต (Rocchetti *et.al.*, 2020) โดยใช้ข้อมูลจำนวนผู้ติดเชื้อและจำนวนผู้เสียชีวิตรายวันซึ่งจะเป็นประโยชน์ในการวางแผนเกี่ยวกับการควบคุมและป้องกันการติดเชื้อโควิด-19 และวางแผนรับมือต่อการแพร่ระบาดของโรคได้อย่างมีประสิทธิภาพ

2. วิธีดำเนินการวิจัย

2.1 ข้อมูลที่ใช้ในการศึกษา

ข้อมูลที่ใช้ในงานวิจัยนี้ คือ จำนวนผู้ติดเชื้อโควิด-19 และจำนวนผู้เสียชีวิตในกรุงเทพมหานครที่ได้รับการยืนยันจากกรมควบคุมโรค กระทรวงสาธารณสุข และข้อมูลการติดเชื้อจากการตรวจเชิงรุกที่ได้รับการยืนยันจากสำนักอนามัย กรุงเทพมหานครตั้งแต่วันที่ 1 พฤศจิกายน พ.ศ. 2564 ถึงวันที่ 31 มกราคม พ.ศ. 2565

2.2 การวิเคราะห์ข้อมูล

2.2.1 คำนวณค่าที่ใช้ในการประมาณขนาดประชากรโดยใช้แนวคิดของ Böhning และคณะ (Böhning *et.al.*, 2020)

$$f_1(t) = \Delta N(t)$$

$$f_2(t) = \Delta N(t-1) - \Delta D(t)$$

$$f_3(t) = \Delta N(t-2) - \Delta D(t-1) - \Delta D(t)$$

$$n^*(t) = f_1(t) + f_2(t) + f_3(t)$$

โดย $\Delta N(t)$ แทน จำนวนผู้ติดเชื้อ ณ วันที่ t

$\Delta N(t-1)$ แทน จำนวนผู้ติดเชื้อ ณ วันที่ $t-1$

$\Delta N(t-2)$ แทน จำนวนผู้ติดเชื้อ ณ วันที่ $t-2$

$\Delta D(t)$ แทน จำนวนผู้เสียชีวิต ณ วันที่ t

$\Delta D(t-1)$ แทน จำนวนผู้เสียชีวิต ณ วันที่ $t-1$

2.2.2 ประเมินจำนวนผู้ติดเชื้อซ่อนเร้น ณ วันที่ t

ตัวประมาณค่าภายใต้การแจกแจงเรขาคณิต

(Niwitpong *et.al.*, 2013, Böhning *et.al.*, 2020)

$$\hat{f}_0^C(t) = \frac{f_1^2(t)}{f_2(t)}, \quad t = t_0 + 1, \dots, t_m \quad (4)$$

ตัวประมาณค่าภายใต้การแจกแจงเรขาคณิต

(Rocchetti *et.al.*, 2020)

$$\hat{f}_0^{UB}(t) = n^*(t) \left(\frac{\hat{\pi}_0^{UB}(t)}{1 - \hat{\pi}_0^{UB}(t)} \right), \quad t = t_0 + 2, \dots, t_m \quad (5)$$

$$\text{เมื่อ } \hat{\pi}_0^{UB}(t) = \frac{\frac{f_1^2(t)}{f_2(t)}}{f_1(t) + f_2(t) + \frac{f_1^2(t)}{f_2(t)}}$$

2.2.3 ประเมินจำนวนผู้ติดเชื้อและความแปรปรวนของจำนวนผู้ติดเชื้อ ณ วันที่ t

ตัวประมาณค่าภายใต้การแจกแจงเรขาคณิต

(Niwitpong *et.al.*, 2013, Böhning *et.al.*, 2020)

$$\hat{N}_C(t) = \Delta N(t) + \hat{f}_0^C(t) \quad (6)$$

$$\text{Var}(\hat{N}_C(t)) = \frac{f_1^4(t)}{f_2^3(t)} + \frac{4f_1^3(t)}{f_2^2(t)} + \frac{f_1^2(t)}{f_2(t)} \quad (7)$$

ตัวประมาณค่าภายใต้การแจกแจงเรขาคณิต

(Rocchetti *et.al.*, 2020)

$$\hat{N}_{UB}(t) = n^*(t) + \hat{f}_0^{UB}(t) \quad (8)$$

2.2.4 ประเมินจำนวนผู้ติดเชื้อและความแปรปรวนของจำนวนผู้ติดเชื้อในช่วงที่ศึกษา

ตัวประมาณค่าภายใต้การแจกแจงเรขาคณิต

(Niwitpong *et.al.*, 2013, Böhning *et.al.*, 2020)

$$\hat{N}_C = \sum_{t=t_0+1}^{t_m} \hat{N}_C(t) \quad (9)$$

$$\text{Var}(\hat{N}_C) = \sum_{t=t_0+1}^{t_m} \text{Var}(\hat{N}_C(t)) \quad (10)$$

ตัวประมาณค่าภายใต้การแจกแจงเรขาคณิต (Rocchetti *et.al.*, 2020)

$$\hat{N}_{UB} = \sum_{t=t_0+2}^{t_m} \hat{N}_{UB}(t) \quad (11)$$

2.2.5 ช่วงความเชื่อมั่นที่ 95% ของจำนวนผู้ติดเชื้อในช่วงที่ศึกษา

เนื่องจากในปี พ.ศ. 2556 Niwitpong และคณะ (Niwitpong *et.al.*, 2013) เสนอตัวประมาณค่าความแปรปรวนของตัวประมาณค่าภายใต้การแจกแจงเรขาคณิต จึงสามารถหาช่วงความเชื่อมั่นโดยวิธีวิลด์ (Pawitan, 2001) ซึ่งอาศัยทฤษฎีบทขีดจำกัดส่วนกลาง (central limit theorem) จะได้ว่า

$$\hat{N}_C \pm z_{\frac{\alpha}{2}} \sqrt{\text{Var}(\hat{N}_C)} \quad (12)$$

โดยที่ $\hat{N}_C$ คือ จำนวนผู้ติดเชื้อในช่วงที่ศึกษาที่ประมาณโดยตัวประมาณค่าภายใต้การแจกแจงเรขาคณิต $z_{\frac{\alpha}{2}}$ คือ

ค่าควอนไทล์ที่ $\frac{\alpha}{2}$ ของการแจกแจงปกติมาตรฐาน

$\text{Var}(\hat{N}_C)$ คือ ค่าประมาณความแปรปรวนของจำนวนผู้ติดเชื้อในช่วงที่ศึกษาที่ประมาณโดยตัวประมาณค่าภายใต้การแจกแจงเรขาคณิต

ดังนั้นช่วงความเชื่อมั่นที่ 95% ของตัวประมาณค่าภายใต้การแจกแจงเรขาคณิตโดยวิธีวิลด์ คือ

$$\hat{N}_C \pm 1.96 \sqrt{\text{Var}(\hat{N}_C)} \quad (13)$$

สำหรับช่วงความเชื่อมั่นของตัวประมาณค่าภายใต้การแจกแจงเรขาคณิต จะใช้วิธีบูตสแตรป์ (bootstrap method) (Efron, 1992) โดยทำการสุ่มตัวอย่างขนาด t_m แบบคืนที่ (resampling with replacement) จำนวน 1,000 ชุด จากค่าสังเกต $\Delta N(1), \Delta N(2), \Delta N(3), \dots, \Delta N(t_m)$ ได้ตัวอย่างชุดใหม่จำนวน 1,000 ชุด คือ $\Delta N(1)^{(b)}, \Delta N(2)^{(b)}, \Delta N(3)^{(b)}, \dots, \Delta N(t_m)^{(b)}$ เมื่อกำหนดให้ $b = 1, 2, \dots, 1000$ หลังจากนั้นคำนวณจำนวนผู้ติดเชื้อในช่วงที่ศึกษาโดยตัวประมาณค่าภายใต้การแจกแจงเรขาคณิตจากสมการที่ 11 ของตัวอย่างแต่ละ

ตารางที่ 1 จำนวนผู้ติดเชื้อที่มีการรายงาน จำนวนผู้ติดเชื้อจากตัวประมาณชาวภายใต้การแจกแจงเรขาคณิต และอัตราส่วนของจำนวนผู้ติดเชื้อจากการประมาณกับจำนวนผู้ติดเชื้อที่มีการรายงานของกรุงเทพมหานครในเดือนพฤศจิกายน พ.ศ. 2564 ถึง มกราคม พ.ศ. 2565

เดือน	จำนวนผู้ติดเชื้อที่มีการรายงาน (n)	$\hat{f}_0$	$\hat{N}_C$	ช่วงความเชื่อมั่น 95%	$\frac{\hat{N}_C}{n}$
พฤศจิกายน	22,783	23,351	46,134	(27,773, 46,496)	2.0249
ธันวาคม	17,635	17,792	35,427	(21,354, 35,745)	2.0089
มกราคม	28,044	30,039	58,083	(35,565, 58,488)	2.0711

ตารางที่ 2 จำนวนผู้ติดเชื้อที่มีการรายงาน จำนวนผู้ติดเชื้อจากตัวประมาณขอบเขตบนภายใต้การแจกแจงเรขาคณิต และอัตราส่วนของจำนวนผู้ติดเชื้อจากการประมาณกับจำนวนผู้ติดเชื้อที่มีการรายงานของกรุงเทพมหานครในเดือน พฤศจิกายน พ.ศ. 2564 ถึง มกราคม พ.ศ. 2565

เดือน	จำนวนผู้ติดเชื้อที่มีการรายงาน (n)	$\hat{f}_0^{UB}$	$\hat{N}_{UB}$	ช่วงความเชื่อมั่น 95%	$\frac{\hat{N}_{UB}}{n}$
พฤศจิกายน	22,783	77,298	99,258	(94,794, 103,774)	4.3567
ธันวาคม	17,635	61,157	76,443	(71,053, 85,834)	4.3347
มกราคม	28,044	97,741	125,377	(106,902, 143,647)	4.4707

ตารางที่ 3 เปรียบเทียบอัตราการติดเชื้อจากตัวประมาณและอัตราการติดเชื้อจากการตรวจเชิงรุกของกรุงเทพมหานครที่มีการรายงานในเดือนพฤศจิกายน พ.ศ. 2564 ถึง มกราคม พ.ศ. 2565

เดือน	จำนวนผู้เข้ารับ การตรวจเชิงรุก	จำนวนผู้ติดเชื้อ จากการตรวจเชิงรุก	$\hat{\pi}_{Active}$ (%)	$\hat{\pi}_C$ (%) (ช่วงความเชื่อมั่น 95%)	$\hat{\pi}_{UB}$ (%) (ช่วงความเชื่อมั่น 95%)
พฤศจิกายน	29,729	456	1.53	0.77 (0.7577, 0.7827)	1.66 (1.5825, 1.7325)
ธันวาคม	38,564	506	1.30	0.59 (0.5806, 0.6023)	1.28 (1.1862, 1.4330)
มกราคม	42,242	763	1.81	0.97 (0.9550, 0.9844)	2.09 (1.7847, 2.3981)

ชุดจะได้ $\hat{N}_{UB}^1, \hat{N}_{UB}^2, \hat{N}_{UB}^3, \dots, \hat{N}_{UB}^{1,000}$ นำค่าที่ได้มาจัดเรียงจากน้อยไปมาก จากนั้นหาค่าตำแหน่งเปอร์เซ็นต์ไทล์ที่ 5 กำหนดให้เป็น $\hat{N}_{UB}^L$ และหาค่าตำแหน่งเปอร์เซ็นต์ไทล์ที่ 95 กำหนดให้เป็น $\hat{N}_{UB}^U$ จะได้ช่วงความเชื่อมั่นที่ 95% ของจำนวนผู้ติดเชื้อในช่วงที่ศึกษาด้วยวิธีบูตสเตรป คือ $[\hat{N}_{UB}^L, \hat{N}_{UB}^U]$

2.3 ประเมินประสิทธิภาพการประมาณจำนวนผู้ติดเชื้อ

ประเมินประสิทธิภาพการประมาณจำนวนผู้ติดเชื้อโดยคำนวณอัตราการติดเชื้อจากตัวประมาณทั้ง 2 เทียบกับอัตราการติดเชื้อจากการตรวจเชิงรุกของกรุงเทพมหานครที่ได้รับการยืนยันจากสำนักอนามัย กรุงเทพมหานคร ตั้งแต่วันที่ 1 พฤศจิกายน พ.ศ. 2564 ถึง วันที่ 31 มกราคม พ.ศ. 2565

$$\hat{\pi}_C = \frac{\text{จำนวนผู้ติดเชื้อจากตัวประมาณชาว}}{\text{จำนวนประชากรในกรุงเทพมหานคร}} \times 100\%$$

$$\hat{\pi}_{UB} = \frac{\text{จำนวนผู้ติดเชื้อจากตัวประมาณขอบเขตบน}}{\text{จำนวนประชากรในกรุงเทพมหานคร}} \times 100\%$$

$$\hat{\pi}_{Active} = \frac{\text{จำนวนผู้ติดเชื้อจากการตรวจเชิงรุก}}{\text{จำนวนผู้เข้ารับการตรวจเชิงรุก}} \times 100\%$$

โดยที่ $\hat{\pi}_C, \hat{\pi}_{UB}, \hat{\pi}_{Active}$ คือ ค่าประมาณของอัตราการติดเชื้อโควิด-19 โดยตัวประมาณชาวภายใต้การแจกแจงเรขาคณิต ตัวประมาณขอบเขตบนภายใต้การแจกแจง

เรขาคณิต และข้อมูลการติดเชื้อจากการตรวจเชิงรุกตามลำดับ หากค่าประมาณของอัตราการติดเชื้อโควิด-19 โดยตัวประมาณชาวภายใต้การแจกแจงเรขาคณิต และตัวประมาณขอบเขตบนภายใต้การแจกแจงเรขาคณิตใกล้เคียงกับอัตราการติดเชื้อจากการตรวจเชิงรุก แสดงว่าตัวประมาณนั้นมีประสิทธิภาพดี

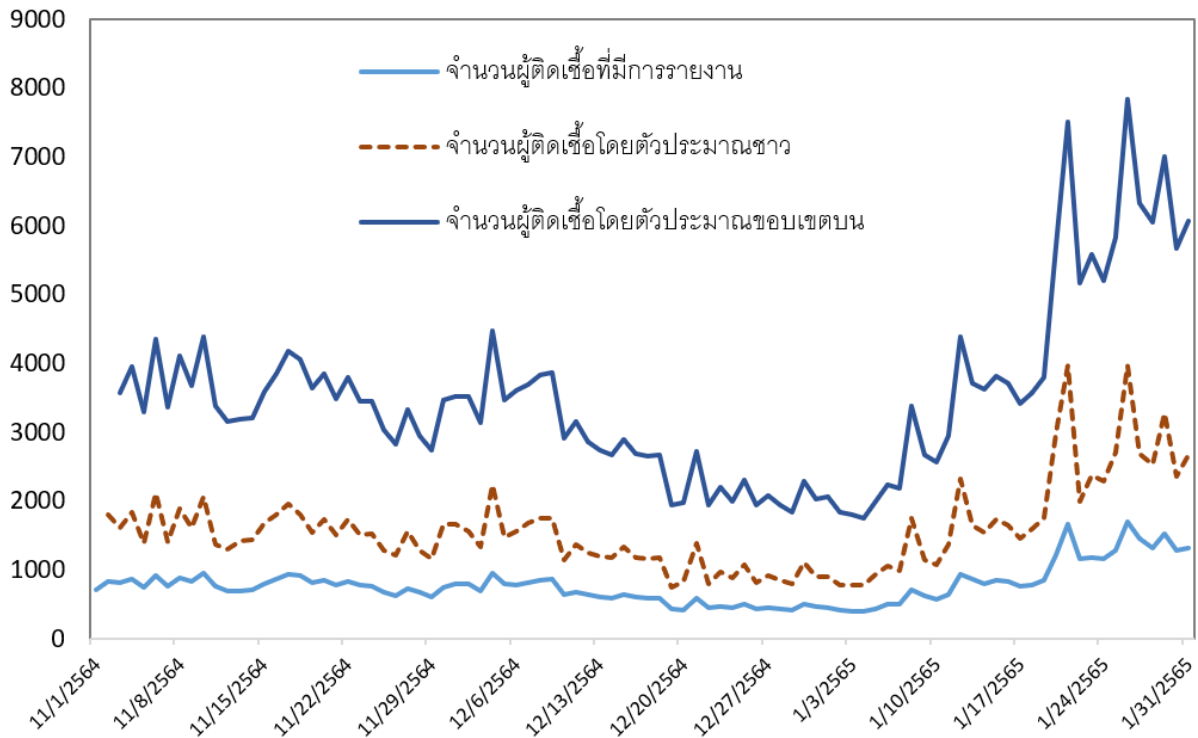
4. ผลการศึกษา

4.1 ผลการประมาณจำนวนผู้ติดเชื้อโควิด-19

ผลการประมาณจำนวนผู้ติดเชื้อโควิด-19 รายเดือนตั้งแต่วันที่ 1 พฤศจิกายน พ.ศ. 2564 ถึง วันที่ 31 มกราคม พ.ศ. 2565 ของกรุงเทพมหานครแสดงได้ดังตารางที่ 1 และ 2

จากข้อมูลจำนวนผู้ติดเชื้อของกรุงเทพมหานครที่รายงานโดยกรมควบคุมโรค พบว่าในเดือนพฤศจิกายน ธันวาคม และมกราคม มีจำนวนผู้ติดเชื้อ 22,783, 17,635 และ 28,044 ราย ตามลำดับ

ตารางที่ 1 แสดงผลการประมาณจำนวนผู้ติดเชื้อโดยตัวประมาณชาวภายใต้การแจกแจงเรขาคณิต พบว่าในเดือนพฤศจิกายน มีจำนวนผู้ติดเชื้อ 46,134 ราย หรือมีจำนวนผู้ติดเชื้อ 27,773 ถึง 46,496 ราย ที่ระดับความเชื่อมั่น 95% ในเดือนธันวาคม มีจำนวนผู้ติดเชื้อ 35,427 ราย หรือมีจำนวนผู้ติดเชื้อ 21,354 ถึง 35,745



รูปที่ 1 จำนวนติดเชื้อโควิด-19 รายวันจากการรายงาน จำนวนผู้ติดเชื้อรายวันจากตัวประมาณชาวภายใต้การแจกแจงเรขาคณิตและตัวประมาณขอบเขตบนภายใต้การแจกแจงเรขาคณิต ในเดือนพฤศจิกายน 2564 พ.ศ.2564 ถึง มกราคม พ.ศ.2565

ราย ที่ระดับความเชื่อมั่น 95% และในเดือนมกราคม มีจำนวนผู้ติดเชื้อ 58,083 ราย หรือมีจำนวนผู้ติดเชื้อ 35,565 ถึง 58,488 ราย ที่ระดับความเชื่อมั่น 95% และยังพบว่าในเดือนพฤศจิกายน ธันวาคม และมกราคม จำนวนผู้ติดเชื้อโดยตัวประมาณชาวภายใต้การแจกแจงเรขาคณิตคิดเป็น 2.0249, 2.0089, 2.0711 เท่าของจำนวนผู้ติดเชื้อที่มีการรายงานโดยกรมควบคุมโรค

ตารางที่ 2 แสดงผลการประมาณจำนวนผู้ติดเชื้อโดยตัวประมาณขอบเขตบนภายใต้การแจกแจงเรขาคณิตพบว่าในเดือนพฤศจิกายน มีจำนวนผู้ติดเชื้อ 99,258 ราย หรือมีจำนวนผู้ติดเชื้อ 94,794 ถึง 103,774 ราย ที่ระดับความเชื่อมั่น 95% ในเดือนธันวาคม มีจำนวนผู้ติดเชื้อ 76,443 ราย หรือมีจำนวนผู้ติดเชื้อ 71,053 ถึง 85,834 ราย ที่ระดับความเชื่อมั่น 95% และในเดือนมกราคม มีจำนวนผู้ติดเชื้อ 125,377 ราย หรือมีจำนวนผู้ติดเชื้อ 106,902 ถึง 143,647 ราย ที่ระดับความเชื่อมั่น 95% และยังพบว่าในเดือนพฤศจิกายน ธันวาคม และมกราคม จำนวนผู้ติดเชื้อโดยตัวประมาณขอบเขตบนภายใต้การแจกแจงเรขาคณิตคิดเป็น 4.3567, 4.3347, 4.4707 เท่าของจำนวนผู้ติดเชื้อที่มีการรายงานโดยกรมควบคุมโรค

4.2 ผลประเมินประสิทธิภาพการประมาณจำนวนผู้ติดเชื้อโควิด-19

ข้อมูลจำนวนผู้เข้ารับการรักษาเชิงรุกและจำนวนผู้ติดเชื้อจากการตรวจเชิงรุกของกรุงเทพมหานครได้รับการยืนยันจากสำนักอนามัย กรุงเทพมหานคร ตั้งแต่วันที่ 1 พฤศจิกายน พ.ศ. 2564 ถึง วันที่ 31 มกราคม พ.ศ. 2565 นำมาคำนวณหาอัตราการติดเชื้อได้ดังตารางที่ 3

ตารางที่ 3 พบว่าในเดือนพฤศจิกายน มีผู้ติดเชื้อจากการตรวจเชิงรุก 456 ราย จากผู้เข้ารับการรักษาทั้งสิ้น 29,729 ราย ซึ่งคิดเป็นอัตราการติดเชื้อร้อยละ 1.53 ในเดือนธันวาคมมีผู้ติดเชื้อจากการตรวจเชิงรุก 506 ราย จากผู้เข้ารับการรักษาทั้งสิ้น 38,564 ราย ซึ่งคิดเป็นอัตราการติดเชื้อร้อยละ 1.30 และในเดือนมกราคมมีผู้ติดเชื้อจากการตรวจเชิงรุก 763 ราย จากผู้เข้ารับการรักษาทั้งสิ้น 42,242 ราย ซึ่งคิดเป็นอัตราการติดเชื้อร้อยละ 1.81

จากการประมาณจำนวนผู้ติดเชื้อโดยตัวประมาณชาวภายใต้การแจกแจงเรขาคณิตในเดือนพฤศจิกายน คำนวณอัตราการติดเชื้อคิดเป็นร้อยละ 0.77 หรือมีอัตราการติดเชื้อคิดเป็นร้อยละ 0.7577 ถึง 0.7827 ที่ระดับความเชื่อมั่น 95% ในเดือนธันวาคม คำนวณอัตราการติดเชื้อคิดเป็นร้อยละ 0.59 หรือมีอัตราการติดเชื้อคิดเป็นร้อยละ 0.5806 ถึง 0.6023 ที่ระดับความเชื่อมั่น 95% และในเดือนมกราคม คำนวณอัตราการติดเชื้อคิดเป็นร้อยละ

0.97 หรือมีอัตราการติดเชื้อคิดเป็นร้อยละ 0.9550 ถึง 0.9844 ที่ระดับความเชื่อมั่น 95%

จากการประมาณจำนวนผู้ติดเชื้อโดยตัวประมาณขอบเขตบนภายใต้การแจกแจงเรขาคณิตในเดือนพฤศจิกายน คำนวณอัตราการติดเชื้อคิดเป็นร้อยละ 1.66 หรือมีอัตราการติดเชื้อคิดเป็นร้อยละ 1.5825 ถึง 1.7325 ที่ระดับความเชื่อมั่น 95% ในเดือนธันวาคม คำนวณอัตราการติดเชื้อคิดเป็นร้อยละ 1.28 หรือมีอัตราการติดเชื้อคิดเป็นร้อยละ 1.1862 ถึง 1.4330 ที่ระดับความเชื่อมั่น 95% และในเดือนมกราคม คำนวณอัตราการติดเชื้อคิดเป็นร้อยละ 2.09 หรือมีอัตราการติดเชื้อคิดเป็นร้อยละ 1.7847 ถึง 2.3981 ที่ระดับความเชื่อมั่น 95%

จากการเปรียบเทียบอัตราการติดเชื้อจากการตรวจเชิงรุกและอัตราการติดเชื้อโดยตัวประมาณทั้งสอง พบว่าอัตราการติดเชื้อโดยตัวประมาณขอบเขตบนภายใต้การแจกแจงเรขาคณิตมีค่าใกล้เคียงกับอัตราการติดเชื้อจากการตรวจเชิงรุกมากกว่าอัตราการติดเชื้อโดยตัวประมาณชาวภายใต้การแจกแจงเรขาคณิต อีกทั้งช่วงความเชื่อมั่นที่ 95% ของอัตราการติดเชื้อโดยตัวประมาณขอบเขตบนภายใต้การแจกแจงเรขาคณิตครอบคลุมอัตราการติดเชื้อจากการตรวจเชิงรุก ดังนั้นตัวประมาณขอบเขตบนภายใต้การแจกแจงเรขาคณิตจึงเหมาะสมที่จะใช้ในการประมาณขนาดประชากรผู้ติดเชื้อโควิด-19 ในกรุงเทพมหานคร

5. อภิปรายและสรุปผล

การประมาณจำนวนประชากรผู้ติดเชื้อโควิด-19 ในกรุงเทพมหานคร โดยตัวประมาณขนาดประชากรชาวภายใต้การแจกแจงเรขาคณิตซึ่งเป็นตัวประมาณขอบเขตล่าง พบว่าในเดือนพฤศจิกายน ธันวาคม พ.ศ. 2564 และ มกราคม พ.ศ. 2565 จะมีจำนวนประชากรผู้ติดเชื้อโควิด-19 อย่างน้อยประมาณ 2 เท่าของจำนวนผู้ติดเชื้อที่มีการรายงานโดยกรมควบคุมโรค และจากตัวประมาณขอบเขตบนภายใต้การแจกแจงเรขาคณิตกล่าวได้ว่าจะมีจำนวนประชากรผู้ติดเชื้อโควิด-19 ไม่เกิน 4.5 เท่าของจำนวนผู้ติดเชื้อที่มีการรายงานโดยกรมควบคุมโรค และเมื่อทำการประเมินประสิทธิภาพของตัวประมาณทั้งสองโดยการคำนวณอัตราการติดเชื้อจากตัวประมาณทั้ง 2 เทียบกับอัตราการติดเชื้อจากการตรวจเชิงรุกของกรุงเทพมหานครที่มีการรายงานโดยสำนักงานผดุงสุขภาพกรุงเทพมหานคร ในช่วงเดือนพฤศจิกายน ธันวาคม และ มกราคม พบว่าอัตราการติดเชื้อจากตัวประมาณขอบเขตบนให้ค่าประมาณใกล้เคียงกับอัตราการติดเชื้อจากการ

ตรวจเชิงรุกมากกว่า และช่วงความเชื่อมั่น 95% ของอัตราการติดเชื้อโดยตัวประมาณขอบเขตบนภายใต้การแจกแจงเรขาคณิตยังครอบคลุมอัตราการติดเชื้อจากการตรวจเชิงรุก จึงกล่าวได้ว่าตัวประมาณขอบเขตบนภายใต้การแจกแจงเรขาคณิตมีประสิทธิภาพในการประมาณจำนวนประชากรผู้ติดเชื้อโควิด-19 ในกรุงเทพมหานคร

กิตติกรรมประกาศ

งานวิจัยนี้เป็นผลงานวิจัยที่ได้รับทุนสนับสนุนการวิจัยประเภทนักวิจัยรุ่นใหม่ จากกองทุนวิจัยคณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์ ตามเลขที่สัญญา 2/2565 ขอขอบคุณ ผศ.ดร.รัตน เลิศสุวรรณศรี ที่ปรึกษาโครงการวิจัย ที่ได้กรุณาให้คำแนะนำซึ่งเป็นประโยชน์อย่างยิ่งต่องานวิจัยนี้

เอกสารอ้างอิง

- Alfö, M., Böhning, D., & Rocchetti, I. (2021). Upper bound estimators of the population size based on ordinal models for capture-recapture experiments. *Biometrics*, 77(1), 237–248.
- Böhning, D., Rocchetti, I., Maruotti, A., & Holling, H. (2020). Estimating the undetected infections in the Covid-19 outbreak by harnessing capture-recapture methods. *International Journal of Infectious Diseases*, 97, 197–201.
- Efron, B., & LePage, R. (1992). *Introduction to bootstrap*. (pp. 3–10). Wiley & Sons.
- Lanumtaeng, K., & Böhning, 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.
- Lerdsuwansri, R., Sangnawakij, P., Böhning, D., Sansilapin, C., Chaifoo, W., Polonsky, J. A., & Del Rio Vilas, V. J. (2022). Sensitivity of contact-tracing for COVID-19 in Thailand: a capture-recapture application. *BMC Infectious Diseases*, 22(1), 1–10.
- Nguyen, L. T., Patel, S., Nguyen, N. T., Gia, H. H., Raymond, H. F., Hoang, V. T. H., & Abdul-Quader, A. S. (2021). Population Size Estimation of Female Sex Workers in Hai Phong, Vietnam: Use of Three Source Capture-Recapture Method. *Journal of Epidemiology and Global Health*, 11(2), 194.
- Niwitpong, S. A., Böhning, D., van der Heijden, P. G., & Holling, H. (2013). Capture-recapture estimation based upon the geometric distribution allowing for heterogeneity. *Metrika*, 76(4), 495–519.
- Pawitan, Y. (2001). *In all likelihood: statistical modelling and inference using likelihood*. Oxford University Press.
- Pijitratana, J., Lerdsuwansri, R., & Böhning, D. (2020). Profile Confidence Interval for Population Size Based on Poisson-Normal Mixture Models. *Proceedings of International Conference in Mathematics and Applications Mahidol University (ICMA-MU2020)* (pp. 119–128).
- Rocchetti, I., Böhning, D., Holling, H., & Maruotti, A. (2020). Estimating the size of undetected cases of the COVID-19 outbreak in Europe: an upper bound estimator. *Epidemiologic Methods*, 9(s1).
- Van Der Heijden, P. G., Bustami, R., Cruyff, M. J., Engbersen, G., & Van Houwelingen, H. C. (2003). Point and interval estimation of the population size using the truncated Poisson regression model. *Statistical Modelling*, 3(4), 305–322.