

การเปรียบเทียบประสิทธิภาพวิธีประมาณค่าเฉลี่ยภายใต้แผนการสุ่มตัวอย่างกลุ่มปรับเมื่อหน่วยตัวอย่างขั้นต้นสุ่ม ด้วยวิธีการแบบง่าย

วิไลวรรณ รัตนกุล¹

¹หลักสูตรวิทยาการข้อมูล คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏอุดรดิตถ์

บทคัดย่อ(ภาษาไทย)

การศึกษาครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพวิธีประมาณค่าเฉลี่ยที่ดัดแปลงมาจากตัวประมาณ Horvitz-Thompson ภายใต้แผนการสุ่มตัวอย่างแบบกลุ่มปรับ โดยวิธีการประมาณค่าอย่างง่าย วิธีการประมาณค่าแบบอัตราส่วน และวิธีการประมาณค่าแบบอัตราส่วนที่ปรับปรุงโดยวิธี Rao-Blackwell ที่หน่วยตัวอย่างใช้แผนการสุ่มตัวอย่างแบบง่าย เกณฑ์ที่ใช้ในการเปรียบเทียบคือ ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย โดยแบ่งพื้นที่การศึกษาออกเป็น 100 หน่วย ประชากรที่ใช้ในการศึกษาสร้างจากกระบวนการปัวซองคลัสเตอร์ที่มีพารามิเตอร์ λ เป็น 30 โดยจำลองประชากรเป็น 2 กรณี ขนาดตัวอย่างที่ใช้ในการศึกษา คือ 3, 6, 9, 12, 16, 20 และ 24 ทำการจำลองสถานการณ์ต่างๆโดยการทำซ้ำ 1,000 ครั้ง ผลการศึกษาพบว่า วิธีการประมาณค่าแบบอัตราส่วนที่ปรับปรุงโดยวิธี Rao-Blackwell เป็นวิธีการที่ดีที่สุด รองลงมาคือวิธีการประมาณค่าแบบอัตราส่วน และวิธีการประมาณค่าอย่างง่าย ตามลำดับทั้ง 2 กรณี

*ผู้เขียนหลัก: wilaiwan.rat@live.ur.ac.th

คำสำคัญ: กระบวนการปัวซองคลัสเตอร์ การสุ่มตัวอย่างแบบกลุ่มปรับ แผนการสุ่มตัวอย่างแบบง่าย การประมาณค่าแบบอัตราส่วน การประมาณค่าแบบอัตราส่วนที่ปรับปรุงโดยวิธี Rao-Blackwell

SCIENCE AND TECHNOLOGY
UTTARADIT RAJABHAT UNIVERSITY

The Efficiency Comparison of Mean Estimator under Adaptive Cluster Sampling with Simple Random Unit Selection

Wilaiwan Rattanakul¹

¹Department of Data Science, Faculty of Science and Technology, Uttaradit Rajabhat University.

Abstract

The objective of this research is to compare the efficiency of mean estimator that mean estimator which have been modified from the Horvitz-Thompson under adaptive cluster sampling when using simple Random unit selection. The simple estimation, ratio estimator and Rao-Blackwell estimator are used for estimation. The mean square error (MSE) is used as a criterion for comparison. The studied areas have been divided into 100 units. The populations are simulated by Poisson cluster process with parameter $\lambda=30$. Simulation of populations in 2 cases and sample size of 3, 6, 9, 12, 16, 20 and 24. In each situation, data are simulated and repeated 1000 time. The result shows Rao-Blackwell estimator is most efficiency in 2 cases and in every sample size.

*Corresponding Author: wilaiwan.rat@live.ur.ac.th

Keywords: Poisson Cluster Process, Adaptive Cluster Sampling, Simple Random, Ratio Estimator, Improved ratio estimator by Rao-Blackwell process

SCIENCE AND TECHNOLOGY
UTTARADIT RAJABHAT UNIVERSITY

1. บทนำ

การเลือกหน่วยตัวอย่างโดยใช้แผนการสุ่มตัวอย่างความน่าจะเป็นขั้นพื้นฐาน(Conventional Probability Sampling) เช่น แผนการสุ่มตัวอย่างแบบง่าย แผนการสุ่มตัวอย่างแบบมีชั้นภูมิ และแผนการสุ่มตัวอย่างแบบมีระบบเป็นต้น กับประชากรที่มีลักษณะที่หายากและอยู่รวมกันเป็นกลุ่มอาจจะทำให้เกิดความเอนเอียง [1] เนื่องจากเมื่อทำการสุ่มตัวอย่างแล้วอาจจะไม่พบสิ่งที่สนใจ หรืออาจจะพบสิ่งที่น่าสนใจเพียงหน่วยเดียวหรือน้อยมาก โดยข้อมูลที่พบไม่เพียงพอในการนำไปคำนวณ ส่งผลให้การประมาณค่าประชากรไม่ตรงกับความเป็นจริง ดังนั้นเพื่อแก้ไขปัญหาดังกล่าวจึงใช้วิธีการสุ่มตัวอย่างแบบกลุ่มปรับ [2] ซึ่งเป็นวิธีการที่ดัดแปลงมาจากแผนการสุ่มตัวอย่างแบบกลุ่ม (Cluster Sampling) เพื่อใช้สำรวจสิ่งที่หายากที่อยู่รวมกันเป็นกลุ่ม เช่น สัตว์ พืช แร่ธาตุ แหล่งฟอสซิล และการศึกษาเกี่ยวกับการระบาดของโรค เป็นต้น โดยเมื่อใดก็ตามที่ทำการสุ่มตัวอย่างแล้วพบสิ่งที่สนใจ ซึ่งเป็นไปได้ที่แหล่งบริเวณที่ติดต่อกันจะพบสิ่งที่สนใจเพิ่มขึ้น ดังนั้นจึงทำการพิจารณาหน่วยบริเวณใกล้เคียงถ้าเป็นไปตามเงื่อนไขก็จะทำการขยายหน่วยถัดไปทางซ้าย ขวา บน และล่าง ทำเช่นนี้ไปเรื่อย ๆ จนกระทั่งไม่พบสิ่งที่สนใจ ข้อมูลที่พบก็จะเพียงพอในการไปคำนวณ โดยการเพิ่มตัวอย่างในหน่วยที่ใกล้เคียง จะมีอัตราการพบสิ่งที่สนใจศึกษาในอัตราที่ค่อนข้างสูง [3] และจากการศึกษางานวิจัยที่เกี่ยวข้องสามารถสรุปได้ดังนี้

Thompson (1990) ศึกษาและเปรียบเทียบประสิทธิภาพของแผนการสุ่มตัวอย่างแบบกลุ่มธรรมดาที่ยังไม่มีการปรับกลุ่มกับแผนการสุ่มตัวอย่างแบบปรับกลุ่มที่หน่วยตัวอย่างใช้วิธีการสุ่มอย่างง่าย เกณฑ์ที่ใช้ในการเปรียบเทียบคือ ค่าความแปรปรวน และค่าความเอนเอียง ผลการวิจัยพบว่า ถ้าใช้ตัวประมาณของแผนการสุ่มตัวอย่างแบบกลุ่มธรรมดาที่ยังไม่มีการปรับกลุ่มนั้นมาประมาณค่าเฉลี่ยสิ่งที่สนใจมีลักษณะหายากและอยู่รวมกันเป็นกลุ่มแล้วนั้น จะทำให้เกิดความเอนเอียงและพบว่าตัวประมาณค่าเฉลี่ยที่ดัดแปลงมาจากตัวประมาณ Horvitz-Thompson มีค่าความแปรปรวนต่ำกว่าตัวประมาณค่าเฉลี่ยที่ดัดแปลงมาจากตัวประมาณ Hansen-Hurwitz และต่อมา Driver & Thompson (2005) ศึกษาและเปรียบเทียบประสิทธิภาพตัวประมาณค่าเฉลี่ยที่ปรับปรุงโดยวิธีการ Rao-Blackwell กับตัวประมาณค่าเฉลี่ยแบบพื้นฐาน ผลการวิจัยพบว่า ตัวประมาณค่าเฉลี่ยอัตราส่วนที่ปรับปรุงโดยวิธีการ Rao-Blackwell จะมีประสิทธิภาพที่ดีกว่าตัวประมาณค่าเฉลี่ยแบบพื้นฐาน Chao et al.(2008) ศึกษาและเปรียบเทียบประสิทธิภาพตัวประมาณค่าเฉลี่ยแบบอัตราส่วนที่ปรับปรุงโดยวิธีการ Rao-Blackwell กับตัวประมาณค่าเฉลี่ยอัตราส่วนแบบพื้นฐานที่ Dryver & Chao (2007) ได้ทำการศึกษาผลการวิจัยพบว่า ตัวประมาณค่าเฉลี่ยแบบอัตราส่วนที่ปรับปรุงโดยวิธีการ Rao-Blackwell จะมีฟังก์ชันของสถิติพอเพียงต่ำที่สุด (Minimal Sufficient Statistic) และมีประสิทธิภาพมากกว่าตัวประมาณค่าเฉลี่ยอัตราส่วนแบบพื้นฐาน

จากการศึกษางานวิจัยดังกล่าวพบว่าเมื่อทำการเปรียบเทียบประสิทธิภาพตัวประมาณค่าเฉลี่ยที่ปรับปรุงโดยวิธีการ Rao-Blackwell กับตัวประมาณค่าเฉลี่ยแบบพื้นฐานพบว่าตัวประมาณค่าเฉลี่ยที่ปรับปรุงโดยวิธีการ Rao-Blackwell ให้ผลการศึกษที่ดีกว่าและการประมาณค่าพารามิเตอร์โดยใช้ตัวแปรช่วยจะทำให้การประมาณค่ามีความคลาดเคลื่อนลดลงดังนั้นผู้วิจัยจึงสนใจเลือกใช้ตัวประมาณที่ดัดแปลงมาจากตัวประมาณ Horvitz-Thompson ซึ่งจากงานวิจัยของ Thompson พบว่าเป็นตัวประมาณที่ให้ค่าความแปรปรวนต่ำที่สุดและนอกจากนั้นผู้วิจัยจะพิจารณาการใช้ตัวแปรช่วยมาประมาณค่าพารามิเตอร์เพื่อจะศึกษาว่าตัวแปรช่วยทำให้การประมาณค่ามีความคลาดเคลื่อนลดลง หรือทำให้การประมาณมีความแม่นยำมากขึ้นซึ่งวิธีการประมาณค่าที่ใช้ตัวแปรช่วยได้แก่วิธีการประมาณค่าเฉลี่ยแบบอัตราส่วนและวิธีการประมาณค่าเฉลี่ยแบบอัตราส่วนที่ปรับปรุงโดยวิธี Rao-Blackwell

2. วิธีการดำเนินการ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพวิธีประมาณค่าเฉลี่ยภายใต้แผนการสุ่มตัวอย่างแบบกลุ่มปรับที่หน่วยตัวอย่างสุ่มด้วยแผนการสุ่มตัวอย่างแบบง่าย ซึ่งประกอบด้วย วิธีประมาณค่าอย่างง่าย (Simple Estimator), วิธีประมาณ

ค่าแบบอัตราส่วน (Ratio Estimator) และวิธีประมาณค่าแบบอัตราส่วนที่ปรับปรุงโดยวิธีการ Rao-Blackwell (Rao-Blackwell Estimator) โดยได้กำหนดขอบเขต และประชากร ดังนี้

1. ประชากรกรณีที่ 1 แบ่งเป็น 2 ลักษณะ
 - 1.1 หน่วยขอบแยกกัน 2 กลุ่ม
 - 1.2 หน่วยขอบแยกกัน 3 กลุ่ม
2. ประชากรกรณีที่ 2 แบ่งเป็น 2 ลักษณะ
 - 2.1 หน่วยขอบรวมกัน 2 กลุ่ม
 - 2.2 หน่วยขอบรวมกัน 3 กลุ่ม
3. ตัวแปรในแต่ละสถานการณ์ประกอบด้วย
 - 3.1 ตัวแปร Y คือ ตัวแปรที่ต้องการศึกษา เป็นข้อมูลเชิงปริมาณ
 - 3.2 ตัวแปร X คือ ตัวแปรช่วยของสิ่งที่ต้องการศึกษา เป็นข้อมูลเชิงปริมาณในพื้นที่ย่อย
 - 3.3 กำหนดความสัมพันธ์ระหว่างตัวแปรเชิงปริมาณ Y กับตัวแปรเชิงปริมาณ X
 - ความสัมพันธ์ระดับต่ำอยู่ในช่วง (0 , 0.4)
 - ความสัมพันธ์ระดับปานกลางอยู่ในช่วง (0.4, 0.7)
 - ความสัมพันธ์ระดับสูงอยู่ในช่วง (0.7, 1.00)
4. กำหนดขนาดพื้นที่ย่อยที่ใช้ในการวิจัยเท่ากับ 3, 6, 9, 12, 16, 20 และ 24

ตัวประมาณค่าที่ใช้ในงานวิจัยมีรายละเอียดดังนี้

2.1 วิธีการประมาณค่าอย่างง่าย (Simple Estimator)

เมื่อประชากรประกอบด้วยหน่วยต่างๆ N หน่วย สำหรับแต่ละหน่วยจะกำหนดพื้นที่ใกล้เคียง ซึ่งประกอบด้วยหน่วยอื่นๆในบริเวณที่รวมหน่วย k ด้วย แล้วกำหนดเกณฑ์ในการรวมหน่วยในพื้นที่ใกล้เคียงเข้าในตัวอย่าง เช่น ถ้าพบว่าค่า y_k สูงกว่าค่าหนึ่ง จะรวมหน่วยพื้นที่นี้เข้าในตัวอย่างเป็นต้น สำหรับแต่ละหน่วยในพื้นที่ใกล้เคียงเข้าในตัวอย่าง เช่น ถ้าพบว่า ค่า y_k สูงกว่าค่าหนึ่ง จะรวมหน่วยพื้นที่นี้เข้าในตัวอย่างเป็นต้น สำหรับแต่ละหน่วยที่เพิ่มขึ้นในตัวอย่างจะทำกระบวนการข้างต้นซ้ำ กล่าวคือ สำหรับหน่วยที่เป็นไปตามเกณฑ์ โดยที่เกณฑ์ในการขยายหน่วยถัดไปที่ใช้ในการศึกษาครั้งนี้คือ $y_i \geq 3$ จึงจะนำหน่วยในพื้นที่ใกล้เคียงเข้าในตัวอย่างด้วย แต่หากไม่เป็นไปตามเกณฑ์ก็ไม่รับหน่วยเพิ่ม [4]

กำหนดสัญลักษณ์

- α_k ความน่าจะเป็นที่หน่วยตัวอย่างที่ i อยู่ในเครือข่ายที่ k
- u_{y_k} เป็นผลรวมของสิ่งที่สนใจ (y) ในเครือข่ายที่ k
- m_k เป็นจำนวนของพื้นที่ย่อยในเครือข่ายที่ k
- v เป็นจำนวนเครือข่ายที่แตกต่างกันในตัวอย่าง
- N เป็นจำนวนของพื้นที่ย่อยทั้งหมด
- n เป็นจำนวนพื้นที่ย่อยตัวอย่าง

ตัวประมาณที่ไม่เอนเอียงของค่าเฉลี่ยตัวอย่างดังสมการที่ 1

$$\bar{y}_{HT}^* = \frac{1}{N} \sum_{k=1}^v \frac{u_{y_k}}{\alpha_k} \quad (1)$$

$$\text{เมื่อ } \alpha_k = 1 - \frac{\binom{N-m_k}{n}}{\binom{N}{n}}$$

โดย $\bar{y}_{HT}^*$ เป็นตัวประมาณที่ไม่เอนเอียงของค่าเฉลี่ยแบบง่ายที่ดัดแปลงมาจากตัวประมาณ Horvitz-Thompson เมื่อหน่วยตัวอย่างสุ่มด้วยแผนการสุ่มตัวอย่างแบบง่าย

ความแปรปรวนของค่าประมาณค่าเฉลี่ยตัวอย่างดังสมการที่ 2

$$V(\bar{y}_{HT}^*) = \frac{1}{N^2} \sum_{k=1}^K \sum_{h=1}^K u_{y_k} u_{y_h} \left(\frac{\alpha_{kh} - \alpha_{ki} \alpha_{hj}}{\alpha_{ki} \alpha_{hj}} \right) \quad (2)$$

$$\text{โดยที่ } \alpha_{kh} = 1 - \left\{ \binom{N-m_k}{n} + \binom{N-m_h}{n} - \binom{N-m_k-m_h}{n} \right\} / \binom{N}{n}$$

α_h ความน่าจะเป็นที่หน่วยตัวอย่างที่ j อยู่ในเครือข่ายที่ h

α_{kh} เป็นความน่าจะเป็นร่วมที่หน่วยตัวอย่างที่ i และ j อยู่ในเครือข่ายที่ k และ h ตามลำดับ

u_{y_h} เป็นผลรวมของสิ่งที่สนใจ (y) ในเครือข่ายที่ h

ค่าประมาณความแปรปรวนของค่าประมาณค่าเฉลี่ยตัวอย่างดังสมการที่ 3

$$v(\bar{y}_{HT}^*) = \frac{1}{N^2} \sum_{k=1}^v \sum_{h=1}^v \frac{u_{y_k} u_{y_h}}{\alpha_{kh}} \left(\frac{\alpha_{kh} - \alpha_k \alpha_h}{\alpha_k \alpha_h} \right) \quad (3)$$

2.2 วิธีการประมาณค่าแบบอัตราส่วน (Ratio Estimator)

Dryver (2007) ได้ศึกษาวิธีการประมาณค่าเฉลี่ยและความแปรปรวนของตัวประมาณค่าเฉลี่ยแบบอัตราส่วนภายใต้แผนการสุ่มตัวอย่างแบบกลุ่มปรับโดยมีรายละเอียดดังนี้ [5]

ประมาณค่าเฉลี่ยแบบอัตราส่วนทั่วไปที่เอนเอียงสำหรับค่าเฉลี่ยประชากร μ ดังสมการที่ 4

$$\bar{y}_{R,HT} = \frac{\bar{y}_{HT}^*}{\bar{x}_{HT}^*} \mu_x = \hat{R}_{ACS} \mu_x \quad (4)$$

$$\text{เมื่อ } \bar{y}_{HT}^* = \frac{1}{N} \sum_{k=1}^v \frac{u_{y_k}}{\alpha_k}, \quad \bar{x}_{HT}^* = \frac{1}{N} \sum_{k=1}^v \frac{u_{x_k}}{\alpha_k} \text{ และ } \mu_x = \frac{1}{N} \sum_{i=1}^N x_i$$

ความแปรปรวนของตัวประมาณค่าเฉลี่ยแบบอัตราส่วนทั่วไปดังสมการที่ 5

$$V(\bar{y}_{R,HT}) = \frac{1}{N^2} \sum_{k=1}^K \sum_{h=1}^K u'_{y_k} u'_{y_h} \left(\frac{\alpha_{kh} - \alpha_k \alpha_h}{\alpha_k \alpha_h} \right) \quad (5)$$

$$\text{เมื่อ } u'_{y_k} = u_{y_k} - R_{ACS} u_{x_k}$$

ค่าประมาณความแปรปรวนของตัวประมาณค่าเฉลี่ยอัตราส่วนตัวอย่างทั่วไปดังสมการที่ 6

$$v(\bar{y}_{R,HT}) = \frac{1}{N^2} \sum_{k=1}^v \sum_{h=1}^v \frac{u'_{y_k} u'_{y_h}}{\alpha_{kh}} \left(\frac{\alpha_{kh} - \alpha_k \alpha_h}{\alpha_k \alpha_h} \right) \quad (6)$$

2.3 วิธีการประมาณค่าแบบอัตราส่วนที่ปรับปรุงโดยวิธีการ Rao - Blackwell (Rao-Blackwell Estimator)

ตัวประมาณอัตราส่วนทั่วไปที่ยังไม่มีการปรับปรุงโดยวิธีการ Rao-Blackwell จะไม่มีฟังก์ชันของสถิติพอเพียงต่ำสุด ซึ่งเป็นคุณสมบัติที่สำคัญทางสถิติ ดังนั้น Chao et al.(2008) ได้ทำการศึกษาตัวประมาณค่าเฉลี่ยแบบอัตราส่วนที่ปรับปรุงโดยวิธีการ Rao-Blackwell เพื่อทำให้ตัวประมาณที่ได้มีสถิติพอเพียงต่ำสุด [6] ซึ่งมีรายละเอียดดังนี้

ตัวประมาณค่าเฉลี่ยแบบอัตราส่วนที่ปรับปรุงด้วยวิธีการ Rao-Blackwell ดังสมการที่ 7

$$\bar{y}_{R,HT}^+ = \frac{\bar{y}_{HT}^+}{\bar{x}_{HT}^+} \mu_x = \frac{\sum_{k=1}^K \frac{u_{y_k}}{\alpha_k} \left(1 - \sum_{i \in \psi_k} e_i \right) + \frac{n}{N} \bar{y}_e}{\sum_{k=1}^K \frac{u_{x_k}}{\alpha_k} \left(1 - \sum_{i \in \psi_k} e_i \right) + \frac{n}{N} \bar{x}_e} \mu_x = \hat{R}_{ACS}^+ \mu_x \quad (7)$$

โดยที่ $\bar{y}_{HT}^+ = \frac{1}{N} \sum_{k=1}^v \frac{m_k \cdot w_{y_k}^+}{\alpha_k}$, $\mu_x = \frac{1}{N} \sum x_i$, $\bar{x}_{HT}^+ = \frac{1}{N} \sum_{k=1}^v \frac{m_k \cdot w_{x_k}^+}{\alpha_k}$

เมื่อ $w_{y_k}^+ = \begin{cases} \bar{y}_e = \frac{\sum_{i \in s} y_i}{e_s} & \text{เมื่อ } \sum_{i \in \psi_k} e_i = 1 \\ w_{y_k} & \text{เมื่อ } \sum_{i \in \psi_k} e_i = 0 \end{cases}$

โดยที่ $w_{y_k} = \frac{1}{m_k} \sum u_{y_k}$ และ $e_i = \begin{cases} 1, & \text{ไม่พบตามเงื่อนไขแต่อยู่บริเวณใกล้เคียงกับส่วนที่เป็นไปตามเงื่อนไข} \\ 0, & \text{อื่น ๆ} \end{cases}$

โดยที่ $\bar{y}_{R,HT}^+$ เป็นตัวประมาณค่าเฉลี่ยแบบอัตราส่วนที่ปรับปรุงด้วยวิธีการ Rao-Blackwell ที่ตัวประมาณดัดแปลงมาจาก Horvitz-Thompson

$\bar{y}_e$ เป็นค่าเฉลี่ยของค่า y สำหรับหน่วยขอบตัวอย่าง

$\bar{x}_e$ เป็นค่าเฉลี่ยของค่า x สำหรับหน่วยขอบตัวอย่าง

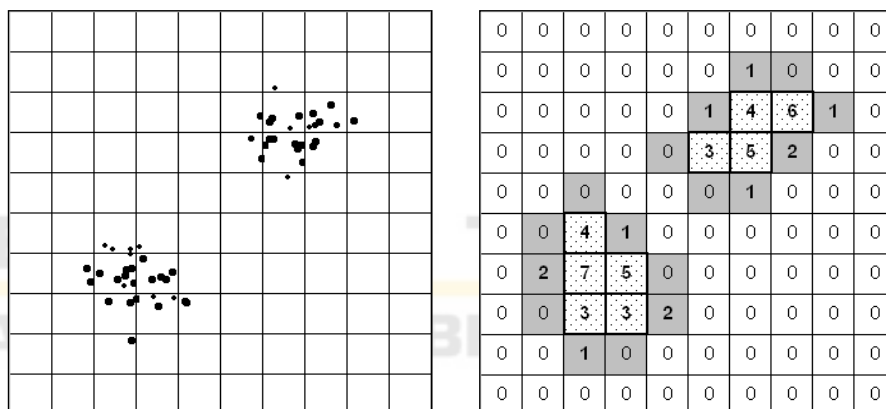
e_s เป็นจำนวนหน่วยขอบตัวอย่างใน s

w_{y_k} เป็นค่าเฉลี่ยของค่า y ในเครือข่ายที่ k

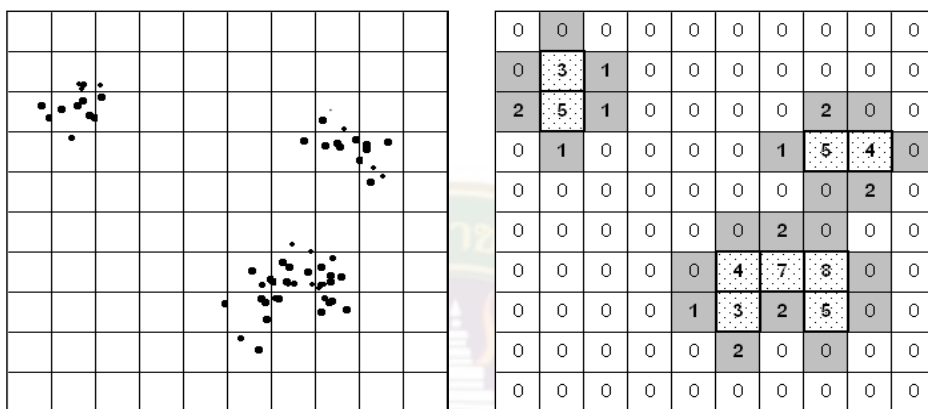
ค่าประมาณความแปรปรวนของตัวประมาณค่าเฉลี่ยแบบอัตราส่วนที่ปรับปรุงด้วยวิธีการ Rao-Blackwell ดังสมการ ที่ 8

$$v(\bar{y}_{R,HT}^+) = v(\bar{y}_{R,HT}) - (\bar{y}_{R,HT}^+ - \bar{y}_{R,HT})^2 \quad (8)$$

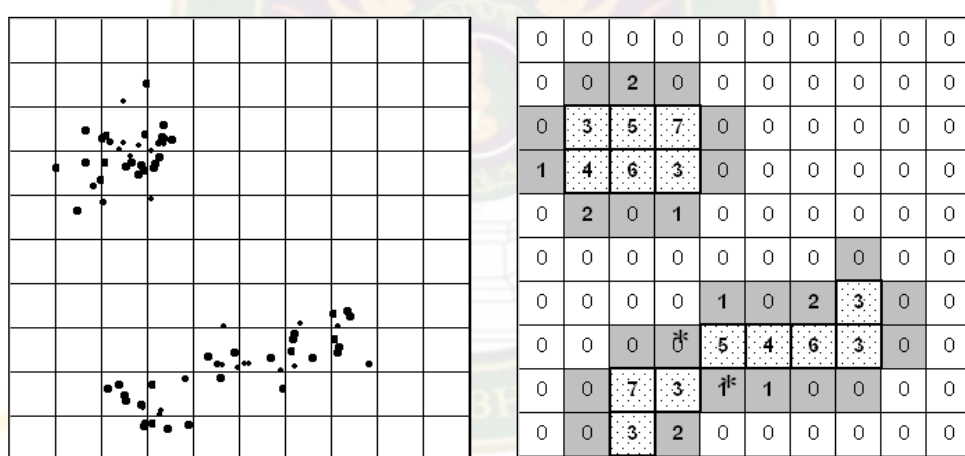
ประชากรที่ใช้ในการศึกษา โดยอาศัยกระบวนการปัวซองคลัสเตอร์ (Poisson Cluster Process) [7] ขนาด 100 หน่วย โดยแบ่งเป็น 2 กลุ่ม



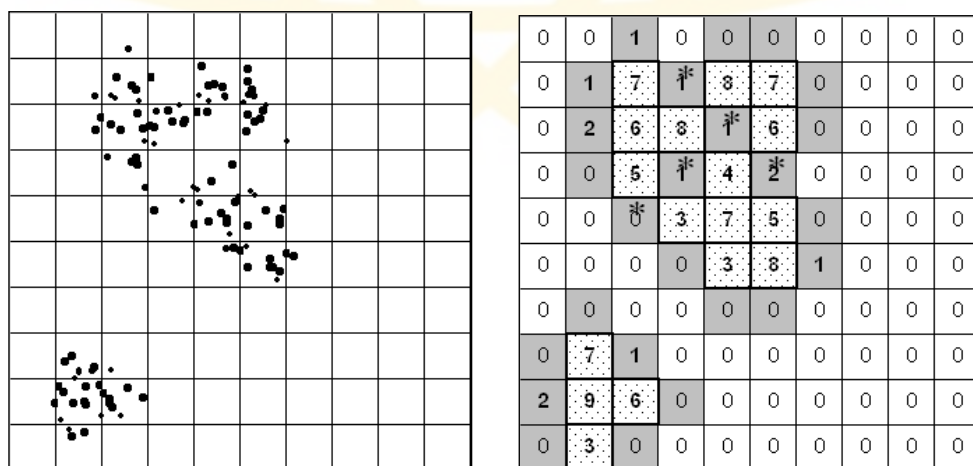
รูปที่ 1 ที่ตั้งและจำนวนของประชากรที่มีลักษณะหายากกรณีที่ 1 เมื่อหน่วยขอบแยกกัน 2 กลุ่ม



รูปที่ 2 ที่ตั้งและจำนวนของประชากรที่มีลักษณะหายากกรณีที่ 1 หน่วยขอบแยกกัน 3 กลุ่ม



รูปที่ 3 ที่ตั้งและจำนวนของประชากรที่มีลักษณะหายากกรณีที่ 2 เมื่อหน่วยขอบร่วมกัน 2 กลุ่ม



รูปที่ 4 ที่ตั้งและจำนวนของประชากรที่มีลักษณะหายากกรณีที่ 2 เมื่อหน่วยขอบร่วมกัน 3 กลุ่ม

3. เนื้อหาหลัก

จากผลการวิจัย เมื่อพิจารณาภายใต้ประชากร 2 กรณีหลักที่แตกต่างกัน โดยแบ่งเป็นประชากรที่มีหน่วยขอบแยกกัน อย่างชัดเจน ซึ่งได้แก่ประชากรที่มีลักษณะหายากกรณีที่ 1 และหน่วยขอบที่ร่วมกันได้แก่ประชากรที่มีลักษณะหายากกรณีที่ 2 ผลที่ได้จากการศึกษาแสดงดังตาราง 1, 2, 3 และ 4 โดยใช้ค่า MSE เป็นเกณฑ์ในการเปรียบเทียบ และใช้สัญลักษณ์แทน

ความหมายของค่าต่าง ๆ ดังนี้ n แทน ขนาดตัวอย่าง SE แทนวิธีการประมาณค่าเฉลี่ยอย่างง่าย RE แทนวิธีการประมาณค่าเฉลี่ยแบบอัตราส่วนและ RB แทนวิธีการประมาณค่าเฉลี่ยแบบอัตราส่วนที่ปรับปรุงโดยวิธีการ Rao-Blackwell

ตารางที่ 1 ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย จำแนกตามระดับความสัมพันธ์ และขนาดพื้นที่ย่อยภายใต้ประชากรที่มีลักษณะหายากกรณีที่ 1 เมื่อหน่วยขอบแยกกัน 2 กลุ่ม

ระดับความสัมพันธ์	n	ตัวประมาณค่าเฉลี่ย		
		SE	RE	RB
ต่ำ	3	1.0061	1.0495	1.0737
	6	0.8904	1.0553	1.0850
	9	0.6921	1.0958	1.1320
	12	0.6295	1.0617	1.1351
	16	0.5238	1.0337	1.1119
	20	0.4859	0.9675	1.0504
	24	0.4610	0.9149	1.0450
ปานกลาง	3	0.9245	0.3995	0.4088
	6	0.7135	0.3487	0.3608
	9	0.6567	0.3503	0.3720
	12	0.5585	0.3292	0.3469
	16	0.4713	0.3150	0.3381
	20	0.4978	0.2938	0.3168
	24	0.4095	0.2556	0.2775
สูง	3	0.3959	0.1644	0.1563
	6	0.3648	0.1514	0.1437
	9	0.3296	0.1369	0.1362
	12	0.2549	0.1148	0.1086
	16	0.2318	0.0911	0.0929
	20	0.2126	0.0793	0.0813
	24	1.1175	0.1914	0.1911

หมายเหตุ:ตัวเลขที่ขีดเส้นใต้ แทนความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุดในแต่ละสถานการณ์

ตารางที่ 2 ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย จำแนกตามระดับความสัมพันธ์ และ ขนาดพื้นที่ย่อยภายใต้ประชากรที่มีลักษณะหายากกรณีที่ 1 เมื่อหน่วยขอบแยกกัน 3 กลุ่ม

ระดับความสัมพันธ์	n	ตัวประมาณค่าเฉลี่ย		
		SE	RE	RB
ต่ำ	3	<u>1.0061</u>	1.0495	1.0737
	6	<u>0.8904</u>	1.0553	1.0850
	9	<u>0.6921</u>	1.0958	1.1320
	12	<u>0.6295</u>	1.0617	1.1351
	16	<u>0.5238</u>	1.0337	1.1119
	20	<u>0.4859</u>	0.9675	1.0504
	24	<u>0.4610</u>	0.9149	1.0450
ปานกลาง	3	0.9245	<u>0.3995</u>	0.4088
	6	0.7135	<u>0.3487</u>	0.3608
	9	0.6567	<u>0.3503</u>	0.3720
	12	0.5585	<u>0.3292</u>	0.3469
	16	0.4713	<u>0.3150</u>	0.3381
	20	0.4978	<u>0.2938</u>	0.3168
สูง	24	0.4095	<u>0.2556</u>	0.2775
	3	1.0736	0.1759	<u>0.1753</u>
	6	0.8536	0.1560	<u>0.1553</u>
	9	0.6793	0.1340	<u>0.1329</u>
	12	0.6127	0.1133	<u>0.1117</u>
	16	0.5400	0.0914	<u>0.0898</u>
	20	1.0736	0.1759	<u>0.1753</u>
	24	0.8536	0.1560	<u>0.1553</u>

หมายเหตุ: ตัวเลขที่ขีดเส้นใต้ แทน ความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุดในแต่ละ สถานการณ์

ตารางที่ 3 แสดงค่าความคลาดเคลื่อนกำลังสองเฉลี่ย จำแนกตามระดับความสัมพันธ์ และ ขนาดพื้นที่ย่อยภายใต้ประชากรที่มีลักษณะหายากกรณีที่ 2 เมื่อหน่วยขอบร่วมกัน 2 กลุ่ม

ระดับความสัมพันธ์	n	ตัวประมาณค่าเฉลี่ย		
		SE	RE	RB
ต่ำ	3	0.9196	0.9612	0.9570
	6	0.7327	0.8802	0.8778
	9	0.6961	0.9084	0.8911
	12	0.5943	0.8228	0.8213
	16	0.5807	0.7306	0.7364
	20	0.5502	0.6815	0.7171
	24	0.5057	0.6479	0.6700
ปานกลาง	3	1.6053	0.3766	0.3719
	6	1.1925	0.3213	0.3137
	9	0.9919	0.2922	0.2821
	12	0.7823	0.2489	0.2367
	16	0.7472	0.2150	0.1998
	20	0.7011	0.1913	0.1724
	24	0.6156	0.1779	0.1622
สูง	3	0.6321	0.1061	0.1032
	6	0.5424	0.0796	0.0758
	9	0.5579	0.0703	0.0670
	12	0.4853	0.0556	0.0513
	16	0.4579	0.0261	0.0225
	20	0.4309	0.0159	0.0129
	24	0.4560	0.0091	0.0071

หมายเหตุ: ตัวเลขที่ขีดเส้นใต้ แทน ความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุดในแต่ละ สถานการณ์

ตารางที่ 4 แสดงค่าความคลาดเคลื่อนกำลังสองเฉลี่ย จำแนกตามระดับความสัมพันธ์ และ ขนาดพื้นที่ย่อยภายใต้ประชากรที่มีลักษณะหายากกรณี 2 เมื่อหน่วยขอบร่วมกัน 3 กลุ่ม

ระดับความสัมพันธ์	n	ตัวประมาณค่าเฉลี่ย		
		SE	RE	RB
ต่ำ	3	1.9965	2.7579	2.8678
	6	1.7513	2.4567	2.6355
	9	1.5581	2.3937	2.5752
	12	1.4358	1.8362	2.0680
	16	1.2489	1.8033	2.0390
	20	1.1624	1.7026	1.9859
	24	1.0945	1.5035	1.8607
ปานกลาง	3	1.3439	0.5671	0.5483
	6	1.0566	0.5005	0.4693
	9	1.1164	0.4933	0.4928
	12	0.9739	0.4517	0.4523
	16	0.9256	0.3543	0.3788
	20	0.8232	0.2821	0.3278
สูง	24	0.8543	0.2419	0.3270
	3	1.8450	0.1830	0.1796
	6	1.4311	0.1614	0.1578
	9	1.2653	0.1108	0.1066
	12	1.2028	0.0786	0.0741
	16	1.1945	0.0506	0.0459
	20	1.0072	0.0389	0.0343
	24	1.0202	0.0394	0.0350

หมายเหตุ: ตัวเลขขีดเส้นใต้ แทน ความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุดในแต่ละสถานการณ์

4. อภิปรายผล

จากการเปรียบเทียบวิธีประมาณค่าเฉลี่ยของทั้ง 2 กรณี พบว่าเมื่อทุกขนาดพื้นที่ย่อย และความสัมพันธ์อยู่ในระดับต่ำ วิธีการประมาณค่าอย่างง่ายเป็นวิธีที่เหมาะสมสำหรับการประมาณค่า เนื่องจากสูตรในการคำนวณหาค่าประมาณพารามิเตอร์ไม่ขึ้นอยู่กับตัวแปรช่วย จึงทำให้วิธีการประมาณค่าอย่างง่ายเป็นวิธีที่เหมาะสม ส่วนวิธีการประมาณค่าแบบอัตราส่วน สูตรในการคำนวณหาค่าประมาณพารามิเตอร์จะขึ้นอยู่กับตัวแปรช่วย โดยเมื่อตัวแปรช่วย และตัวแปรที่สนใจศึกษา มีความสัมพันธ์อยู่ในระดับต่ำ พบว่าค่า $\hat{Ru}_{x_k}$ จะมีค่าน้อยกว่า u_{y_k} ส่งผลให้ค่า w'_k สูงขึ้น จึงทำให้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยมีค่าสูงขึ้น ในขณะที่ความสัมพัทธ์ระหว่างตัวแปรช่วย และตัวแปรที่สนใจศึกษาอยู่ในระดับปานกลาง และสูง พบว่าค่า $\hat{Ru}_{x_k}$ จะมีค่าใกล้เคียงกับค่า u_{y_k} จึงส่งผลให้ w'_k มีค่าน้อย ดังนั้นค่าความคลาดเคลื่อนกำลังสองเฉลี่ยจึงมีค่าน้อยลง และนอกจากนี้ ผลการวิจัยยังพบว่าวิธีประมาณค่าแบบอัตราส่วนที่ปรับปรุงโดยวิธีการ Rao-Blackwell จะให้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำกว่าวิธีการประมาณค่าแบบอัตราส่วน เนื่องจากค่าความแปรปรวนของวิธีประมาณค่าแบบอัตราส่วนที่ปรับปรุงโดยวิธีการ Rao-Blackwell ได้จากค่าความแปรปรวนของวิธีประมาณค่าแบบอัตราส่วนลบด้วยผลต่างกำลังสองของค่าเฉลี่ย จึงทำให้ค่าประมาณที่ได้มีความแปรปรวนต่ำกว่าวิธีการประมาณค่าแบบอัตราส่วน ซึ่งสอดคล้องกับผลการวิจัยของ Chao, Lin และ Chiang (2008) ที่ทำการศึกษา และเปรียบเทียบประสิทธิภาพตัวประมาณค่าเฉลี่ยแบบอัตราส่วนที่ปรับปรุงโดยวิธีการ Rao-Blackwell กับตัวประมาณอัตราส่วนแบบพื้นฐานผลการวิจัยพบว่า ตัวประมาณค่าเฉลี่ยแบบอัตราส่วนที่ปรับปรุงโดยวิธีการ Rao-Blackwell จะมีประสิทธิภาพมากกว่าตัวประมาณค่าเฉลี่ยอัตราส่วนแบบพื้นฐาน เมื่อพิจารณาผลการวิจัยทั้ง 2 กรณี พบว่ากรณีหน่วยขอบรวมกันหรือหน่วยขอบแยกกัน จะให้ผลที่สอดคล้องในทิศทางเดียวกัน เนื่องจากหน่วยขอบไม่ส่งผลต่อวิธีการประมาณค่า ดังนั้นผลการวิจัยที่จะระดับความสัมพันธ์ต่ำ วิธีการประมาณค่าอย่างง่ายจะเป็นวิธีที่เหมาะสมสำหรับการประมาณค่า และเมื่อความสัมพันธ์อยู่ในระดับปานกลาง และสูง วิธีการประมาณค่าแบบอัตราส่วนที่ปรับปรุงโดยวิธีการ Rao-Blackwell เป็นวิธีที่เหมาะสมสำหรับการประมาณค่า ยกเว้นเมื่อหน่วยขอบแยกกัน 2 และ 3 กลุ่ม ที่ระดับความสัมพันธ์ปานกลาง พบว่าวิธีการประมาณค่าแบบอัตราส่วน เป็นวิธีที่เหมาะสมสำหรับการประมาณค่าเนื่องจากการวิจัยนี้ระบุความสัมพันธ์ระหว่างตัวแปรเป็นช่วง ซึ่งอาจมีโอกาที่ความสัมพันธ์ในระดับปานกลางและสูงจะมีค่าใกล้เคียงกัน ดังนั้นจึงส่งผลให้วิธีการประมาณค่าแบบอัตราส่วนมีค่าความคลาดเคลื่อนกำลังสองเฉลี่ยใกล้เคียงกับวิธีการประมาณค่าแบบอัตราส่วนที่ปรับปรุงโดยวิธีการ Rao-Blackwell

นอกจากนี้ยังพบว่าเมื่อขนาดพื้นที่ย่อย และระดับความสัมพันธ์เพิ่มขึ้นความคลาดเคลื่อนกำลังสองเฉลี่ยของวิธีประมาณทั้งสามวิธีมีแนวโน้มลดลงเนื่องจากขนาดตัวอย่างที่มีค่าใกล้เคียงกับประชากรจะทำให้ค่าความเอนเอียงจากตัวอย่างลดน้อยลง จึงส่งผลให้การประมาณค่าทั้ง 3 วิธีมีประสิทธิภาพสูงขึ้น

เอกสารอ้างอิง

- [1] Thompson, S.K. (1990). Adaptive Cluster Sampling. The American Statistician, 44, 1050-1059
- [2] Mena Lao & Arthur Dryver (2022). Adaptive cluster path sampling. Communications in Statistics - Simulation and Computation. Article | Published Online: 27 May 2022
- [3] สุขาดา กิรินทร์. (2542). ทฤษฎีและวิธีการสำรวจตัวอย่าง. กรุงเทพฯ: สำนักพิมพ์ จุฬาลงกรณ์มหาวิทยาลัย.
- [4] Thompson, S.K.(2002). *Sampling* (2nd ed.). New York: Wiley .
- [5] Dryver, A.L., & Chao, C.T. (2007). Ratio estimators in adaptive cluster sampling. *Environmetric* , 18, 607-620.

- [6] Chao, C.T., Lin, F.M., & Chiang, T.C. (2008). Improved Ratio Estimators in Adaptive Cluster Sampling. *The American Statistician*, 3210-3214.
- [7] Cliff, A. D., & Ord, J. K. (1981). *Spatial Processes Models and Applications*. London: Pion Limited.



SCIENCE AND TECHNOLOGY
UTTARADIT RAJABHAT UNIVERSITY