

การปรับปรุงตัวประมาณค่าเฉลี่ยประชากรด้วยข้อมูลลำดับภายใต้การเลือกตัวอย่างสองครั้ง The Improved Estimator of Population Mean with Ranks under Double Sampling

คณิศา โชติจันตัก^{1*} สรรค์ชัย ชิวขุนทด¹ และ ณัทัย สระกบแก้ว²

Kanisa Chodjuntug^{1*} Sanchai Chewkhunthod¹ and Nahathai Srakobkaew²

¹ภาควิชาคณิตศาสตร์ สถิติ และคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยอุบลราชธานี อุบลราชธานี 34190 ประเทศไทย

²กองวิชาคณิตศาสตร์และคอมพิวเตอร์ ส่วนการศึกษา โรงเรียนนายร้อยพระจุลจอมเกล้า นครนายก 26000 ประเทศไทย

¹Department of Mathematics Statistics and Computer, Faculty of Science, Ubon Ratchathani University,
Ubon Ratchathani 34190, Thailand

²Department of Mathematics and Computer Science, Academic Division, Chulachomklao Royal Military Academy,
Nakhon Nayok 26000, Thailand

*E-mail: kanisa.c@ubu.ac.th

Received: Feb 01, 2025

Revised: Mar 09, 2025

Accepted: Mar 13, 2025

บทคัดย่อ

งานวิจัยนี้นำเสนอการพัฒนาตัวประมาณค่าเฉลี่ยประชากรโดยใช้ลำดับของตัวแปรเสริมภายใต้การเลือกตัวอย่างสองครั้ง โดยศึกษาคุณสมบัติบางประการของตัวประมาณ ได้แก่ ความเอนเอียง (Bias: B) และความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean square error: MSE) และเปรียบเทียบประสิทธิภาพของตัวประมาณที่นำเสนอกับตัวประมาณเดิม คือ วิธีค่าเฉลี่ย และวิธีอัตราส่วน การวิเคราะห์เชิงประจักษ์ใช้ข้อมูลฝุ่นละอองขนาดเล็กในเขตกรุงเทพมหานคร ช่วงเดือนมกราคม 2562 ได้ทำการสุ่มตัวอย่างซ้ำ 10,000 รอบ พบว่า ตัวประมาณที่นำเสนอมีประสิทธิภาพสูงกว่าตัวประมาณเดิม โดยมีค่า MSE ต่ำที่สุดในทุกสถานการณ์ที่ทำการศึกษา และมีแนวโน้มลดลงเมื่อขนาดตัวอย่างเพิ่มขึ้น จึงสรุปได้ว่า การใช้ลำดับของตัวแปรเสริมสามารถลดค่า MSE ได้ ทำให้ตัวประมาณมีความแม่นยำมากขึ้น

คำสำคัญ: การเลือกตัวอย่างสองครั้ง ลำดับของตัวแปรเสริม ตัวประมาณค่าเฉลี่ยประชากร ความคลาดเคลื่อนกำลังสองเฉลี่ย

Abstract

This research presents the development of a population mean estimator using the ranked auxiliary variables under double sampling. Some properties of the proposed estimator were studied, including bias (B) and mean square error (MSE). The efficiency of the proposed estimator was also compared with the standard estimator, namely the mean method and the ratio method. For the empirical analysis, the data of small dust particles in Bangkok in January 2019 were used, with 10,000 replicate samplings. It was found that the proposed estimator was more efficient than the standard estimator with the lowest MSE values in all studied scenarios which tended to decrease as the sample size increased. Therefore, it can be concluded that using the ranked auxiliary variables can reduce the MSE value, resulting in a more accurate estimator.

Keywords: Double sampling, Ranked auxiliary variables, Population mean estimator, Mean square error

1. บทนำ

การประมาณค่าเฉลี่ยของประชากรเป็นการวิเคราะห์ข้อมูลทางด้านสถิติที่มักพบบ่อยในงานวิจัยต่าง ๆ เช่น สิ่งแวดล้อม การแพทย์ การตลาด เป็นต้น เนื่องจากข้อมูลที่ต้องการศึกษามักมีขนาดใหญ่และมีข้อจำกัดด้านเวลาในการทำวิจัยจึงไม่สามารถเก็บรวบรวมข้อมูลทั้งหมด การหาค่าเฉลี่ยของประชากรจึงต้องอาศัยการสำรวจตัวอย่างเพื่อเก็บรวบรวมข้อมูลบางส่วนของประชากรในการวิเคราะห์ข้อมูล นักสถิติจึงสนใจพัฒนาการประมาณค่าเฉลี่ยหรือปรับปรุงประสิทธิภาพของตัวประมาณค่าเฉลี่ยให้มีความแม่นยำหรือใกล้เคียงกับค่าเฉลี่ยประชากรมากที่สุด โดยทั่วไปการประมาณค่าเฉลี่ยประชากรของข้อมูลที่กำหนดให้เป็นตัวแปรที่สนใจศึกษา (Study variable) มักนิยมใช้ คือ ตัวประมาณด้วยค่าเฉลี่ย (Mean estimator) ตัวประมาณอัตราส่วน (Ratio estimator) และตัวประมาณถดถอยเชิงเส้น (Linear regression estimator) อยู่ภายใต้การสุ่มตัวอย่างอย่างง่าย (Simple random sampling) โดยตัวประมาณดังกล่าวจำเป็นต้องใช้ค่าเฉลี่ยประชากรของตัวแปรเสริม (Auxiliary variable) ซึ่งในทางปฏิบัติลักษณะการเก็บข้อมูลจากการสำรวจตัวอย่าง ข้อมูลแต่ละหน่วยจะรวบรวมในรูปแบบคู่ลำดับระหว่างตัวแปรที่สนใจศึกษากับตัวแปรเสริม จึงไม่สามารถหาข้อมูลประชากรของตัวแปรเสริมได้ เพื่อแก้ไขข้อจำกัดนี้แผนการสุ่มตัวอย่างด้วยวิธีการเลือกตัวอย่างสองครั้ง (Double sampling) จึงได้รับการพัฒนา โดยตัวประมาณจะมีการใช้ค่าเฉลี่ยตัวอย่างของตัวแปรเสริมในการเลือกตัวอย่างครั้งแรกแทนการใช้ค่าเฉลี่ยประชากรของตัวแปรเสริมนอกจากนี้การใช้วิธีดังกล่าวยังสามารถลดต้นทุนการเก็บข้อมูลได้อย่างมีประสิทธิภาพ เนื่องจากการเก็บข้อมูลทั้งหมดในครั้งเดียวอาจทำให้ต้นทุนค่าใช้จ่ายสูงเกินไปและไม่เหมาะสมในทางปฏิบัติ

สำหรับวิธีการเลือกตัวอย่างสองครั้ง กำหนดให้ Y คือ ตัวแปรที่สนใจศึกษา และ X คือ ตัวแปรเสริม โดย $U = \{(X_1, Y_1), (X_2, Y_2), (X_3, Y_3), \dots, (X_N, Y_N)\}$ คือ เซตของประชากรขนาด N ดังนั้น $\bar{Y}$ คือ ค่าเฉลี่ยประชากรของตัวแปร Y สนใจศึกษา $\bar{X}$ คือ ค่าเฉลี่ยประชากรของตัวแปร X การเลือกตัวอย่างสองครั้งจะมีการพิจารณาเลือกตัวอย่างโดยครั้งแรกจะเลือกตัวอย่างสุ่มอย่างง่ายแบบไม่ใส่คืน (Simple random sampling without replacement) ด้วยจำนวนตัวอย่างขนาด m หน่วย จากประชากรขนาด N หน่วย ครั้งถัดไปหรือครั้งสอง จะเลือกตัวอย่างสุ่มอย่างง่ายแบบไม่ใส่คืนด้วยจำนวนตัวอย่างขนาด n หน่วย จากตัวอย่าง

ขนาด m หน่วย ในครั้งแรก ดังนั้น $\bar{x}_m$ คือ ค่าเฉลี่ยตัวอย่างของตัวแปร X ในครั้งแรก $\bar{y}_n$ คือ ค่าเฉลี่ยตัวอย่างของตัวแปร Y ในครั้งสอง และ $\bar{x}_n$ คือ ค่าเฉลี่ยตัวอย่างของตัวแปร X ในครั้งสอง

ตัวประมาณด้วยค่าเฉลี่ยภายใต้การเลือกตัวอย่างสองครั้ง [1] ดังสมการ (1)

$$\hat{Y}_{Md} = \bar{y}_n \quad (1)$$

ตัวประมาณอัตราส่วนภายใต้การเลือกตัวอย่างสองครั้ง [2] ดังสมการ (2)

$$\hat{Y}_{Rd} = \bar{y}_n \frac{\bar{x}_m}{\bar{x}_n} \quad (2)$$

ตัวประมาณการถดถอยเชิงเส้นภายใต้การเลือกตัวอย่างสองครั้ง [3] ดังสมการ (3)

$$\hat{Y}_{Regd} = \bar{y}_n + \beta(\bar{x}_m - \bar{x}_n) \quad (3)$$

โดยที่

β = ค่าสัมประสิทธิ์การถดถอย

ต่อมา Khare and Rehman [4] ได้เสนอตัวประมาณใหม่ที่ได้พัฒนาจากตัวประมาณการถดถอยเชิงเส้นภายใต้การเลือกตัวอย่างสองครั้ง และปรับปรุงด้วยการใช้ตัวแปรเสริมสองตัวแปร คือ X และ Z ดังสมการ (4)

$$\hat{Y}_{KR} = k_1 \bar{y}_n + k_2 (\bar{x}_m - \bar{x}_n) + k_3 (\bar{z} - \bar{z}_m) \quad (4)$$

โดยที่

k_1, k_2, k_3 คือ ค่าคงที่

$\bar{z}$ = ค่าเฉลี่ยประชากรของตัวแปร Z

$\bar{z}_m$ = ค่าเฉลี่ยของตัวแปร Z ในครั้งแรก

อย่างไรก็ตามตัวประมาณที่เสนอโดย [4] ต้องใช้ตัวแปรเสริมสองตัว คือ X และ Z ร่วมกับตัวแปร Y และยังคงอาศัยค่าเฉลี่ยประชากรของตัวแปร Z ในการปรับปรุงตัวประมาณ จึงกลายเป็นข้อจำกัดในการนำไปประยุกต์ใช้ในทางปฏิบัติ

ปัจจุบันมีนักวิจัยหลายท่าน [5]-[8] ได้ใช้ข้อมูลลำดับของตัวแปร X (Ranked auxiliary variables) ในการปรับปรุงตัวประมาณค่าเฉลี่ย โดยแสดงให้เห็นว่า การใช้ลำดับของ

ตัวแปร X สามารถลดความคลาดเคลื่อนของตัวประมาณได้ ตัวประมาณดังกล่าวจึงมีประสิทธิภาพดีกว่าตัวประมาณด้วยค่าเฉลี่ย และตัวประมาณอัตราส่วน นอกจากนี้การใช้ลำดับของตัวแปรเสริมยังช่วยลดผลกระทบของค่าผิดปกติและเพิ่มความสัมพันธ์ระหว่างตัวแปร X กับตัวแปร Y จึงทำให้การใช้ข้อมูลดังกล่าวช่วยเพิ่มความแม่นยำของตัวประมาณได้

จากการศึกษาที่ผ่านมา Haq et al. [9] ได้เสนอตัวประมาณค่าที่ใช้ค่าเฉลี่ยและการใช้ลำดับของตัวแปร X ภายใต้การเลือกตัวอย่างสุ่มอย่างง่าย ดังสมการ (5)

$$\hat{Y}_{PR} = k_1 \bar{y} + k_2 (\bar{X} - \bar{x}) + k_3 (\bar{R}_x - \bar{r}_x) \quad (5)$$

โดยที่

$\bar{x}$ คือ ค่าเฉลี่ยตัวอย่างของตัวแปร X ภายใต้การเลือก

ตัวอย่างสุ่มอย่างง่าย

$\bar{y}$ คือ ค่าเฉลี่ยตัวอย่างของตัวแปร Y ภายใต้การเลือก

ตัวอย่างสุ่มอย่างง่าย

$\bar{R}_x$ คือ ค่าเฉลี่ยลำดับประชากรของตัวแปร X

$\bar{r}_x$ คือ ค่าเฉลี่ยลำดับตัวอย่างของตัวแปร X ภายใต้การเลือก

ตัวอย่างสุ่มอย่างง่าย

อย่างไรก็ตามการพัฒนาตัวประมาณด้วยการใช้ข้อมูลลำดับของตัวแปร X มักมุ่งเน้นภายใต้สถานการณ์การเลือกตัวอย่างสุ่มอย่างง่ายเป็นหลัก งานวิจัยนี้จึงสนใจพัฒนาตัวประมาณที่เสนอโดย Haq et al. [9] โดยอาศัยแนวคิดของ Khare and Rehman [4] ภายใต้วิธีการเลือกตัวอย่างสองครั้ง เพื่อแก้ไขปัญหากรณีไม่ทราบข้อมูลประชากร โดยใช้ข้อมูลกลุ่มตัวอย่างแทนข้อมูลประชากร และศึกษาคุณสมบัติบางประการของตัวประมาณ ได้แก่ ความเอนเอียง (Bias: B) และความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean square error: MSE) และเปรียบเทียบประสิทธิภาพตัวประมาณค่าที่นำเสนอกับตัวประมาณค่าที่มีอยู่เดิม คือ $\hat{Y}_{Md}$ และ $\hat{Y}_{Rd}$ และใช้ข้อมูลมลพิษทางอากาศจากกรมควบคุมมลพิษในการเปรียบเทียบประสิทธิภาพในงานวิจัยครั้งนี้

2. วิธีศุ อุปกรณ์และวิธีการวิจัย

การวิจัยครั้งนี้ได้มีการเสนอตัวประมาณค่าเฉลี่ยแบบจุดของตัวแปรที่สนใจศึกษา (Y) ภายใต้ขอบเขตการสุ่มตัวอย่างแบบสองครั้ง กรณีใช้ตัวแปรเสริมเพียงหนึ่งตัว (X) และไม่ทราบค่าประชากร โดยใช้โปรแกรม R เป็นเครื่องมือในการทำวิจัย ซึ่งมีรายละเอียดวิธีการทำวิจัยดังนี้

2.1. นำเสนอตัวประมาณ

งานวิจัยนี้พัฒนาตัวประมาณที่เสนอโดย Haq et al. [9] โดยอาศัยแนวคิดของ Khare and Rehman [4] ภายใต้การสุ่มตัวอย่างสองครั้ง จึงปรับปรุงใหม่จากเดิม $\bar{X}$ แทนด้วย $\bar{x}_m$ $\bar{x}$ แทนด้วย $\bar{x}_n$ $\bar{y}$ แทนด้วย $\bar{y}_n$ $\bar{R}_x$ แทนด้วยค่าเฉลี่ยลำดับของตัวอย่างในการเลือกตัวอย่างครั้งแรก ซึ่งใช้สัญลักษณ์ $\bar{r}_x^{(1)}$ และ $\bar{r}_x$ แทนด้วย ค่าเฉลี่ยลำดับของตัวอย่างในการเลือกตัวอย่างครั้งที่สองซึ่งใช้สัญลักษณ์ $\bar{r}_x^{(2)}$ จะได้ตัวประมาณที่นำเสนอ คือ

$$\hat{Y}_{Newd} = k_1 \bar{y}_n + k_2 (\bar{x}_m - \bar{x}_n) + k_3 (\bar{r}_x^{(1)} - \bar{r}_x^{(2)}) \quad (6)$$

2.2. ในทางทฤษฎีศึกษาคุณสมบัติตัวประมาณที่นำเสนอ

คุณสมบัติของตัวประมาณที่นำเสนอสำหรับศึกษาในครั้งนี้ คือ B และ MSE

2.3. เปรียบเทียบประสิทธิภาพของตัวประมาณ

การศึกษาประสิทธิภาพของตัวประมาณ $\hat{Y}_{Newd}$ เปรียบเทียบกับตัวประมาณเดิม คือ $\hat{Y}_{Md}$ และ $\hat{Y}_{Rd}$ โดยใช้ข้อมูลทุติยภูมิจากกรมควบคุมมลพิษของฝุ่นละอองขนาดเล็กตรวจวัดในเขตพื้นที่กรุงเทพมหานคร บริเวณโรงพยาบาลจุฬาลงกรณ์ช่วงเดือน มกราคม 2562 ด้วยขนาดประชากร (N) เท่ากับ 718 ครั้ง [10] กำหนดให้ Y คือ ค่า $PM_{2.5}$ และ X คือ ค่า PM_{10} เนื่องจากค่า $PM_{2.5}$ และค่า PM_{10} มีความสัมพันธ์กันสูง (High correlation)

ขนาดตัวอย่างในการเลือกครั้งแรก จะสุ่มจากประชากรโดยศึกษา 3 ระดับ คือ 10% 20% และ 50% ส่วนขนาดตัวอย่างในครั้งที่สอง จะสุ่มจากตัวอย่างในครั้งแรกโดยศึกษา 2 ระดับ คือ 20% และ 50% ในแต่ละกรณีทำซ้ำ 10,000 รอบ ภายใต้วิธีการเลือกตัวอย่างสองครั้ง โดยใช้โปรแกรม R ในการวิจัยครั้งนี้ และการเปรียบเทียบประสิทธิภาพของตัวประมาณใช้ MSE ของตัวประมาณใด ๆ ($\hat{Y}$) ดังสมการ (7)

$$MSE(\hat{Y}) = \frac{\sum_{j=1}^{10,000} (\hat{Y}_j - \bar{Y})^2}{10,000} \quad (7)$$

การพิจารณาเปรียบเทียบตัวประมาณโดยตัวประมาณที่มีค่า MSE ต่ำที่สุด จะมีประสิทธิภาพดีที่สุด

3. ผลการวิจัยและอภิปรายผลการวิจัย

3.1. การศึกษาทางทฤษฎี

คุณสมบัติของตัวประมาณที่ศึกษาได้แก่ B และ MSE เพื่อให้ง่ายต่อการหาคุณสมบัติดังกล่าวจึงได้มีการใช้เทคนิคการแปลงตัวประมาณให้อยู่ในรูป $e_i : i=1,2,3,4$ [3] รายละเอียดดังนี้
กำหนดให้

$$e_0 = \frac{(\bar{y}_n - \bar{Y})}{\bar{Y}}, \quad e_1 = \frac{(\bar{x}_n - \bar{X})}{\bar{X}}, \quad e_2 = \frac{(\bar{r}_x^{(2)} - \bar{R}_X)}{\bar{R}_X}$$

$$e_3 = \frac{(\bar{x}_m - \bar{X})}{\bar{X}}, \quad e_4 = \frac{(\bar{r}_x^{(1)} - \bar{R}_X)}{\bar{R}_X}, \quad f_1 = \frac{1}{n} - \frac{1}{N},$$

$$f_2 = \frac{1}{m} - \frac{1}{N}, \quad f_3 = f_1 - f_2$$

จะได้

$$E(e_0) = E(e_1) = E(e_2) = E(e_3) = E(e_4) = 0$$

$$E(e_0^2) = f_1 C_Y^2, E(e_1^2) = f_1 C_X^2, E(e_2^2) = f_1 C_R^2$$

$$E(e_3^2) = E(e_1 e_3) = f_2 C_X^2, \quad E(e_4^2) = E(e_2 e_4) = f_2 C_R^2$$

$$E(e_0 e_1) = f_1 \rho_{XY} C_X C_Y, \quad E(e_0 e_2) = f_1 \rho_{YR} C_Y C_R$$

$$E(e_1 e_2) = f_1 \rho_{XR} C_X C_R, \quad E(e_0 e_3) = f_2 \rho_{XY} C_X C_Y, \quad E(e_0 e_4) = f_2 \rho_{YR} C_Y C_R,$$

$$E(e_1 e_4) = E(e_2 e_3) = E(e_3 e_4) = f_2 \rho_{XR} C_X C_R$$

โดยที่

$$\rho_{XY} = \frac{S_{XY}}{S_Y S_X}, \rho_{YR} = \frac{S_{YR}}{S_Y S_R}, \rho_{XR} = \frac{S_{XR}}{S_X S_R}$$

$$C_X = \frac{S_X}{\bar{X}}, C_Y = \frac{S_Y}{\bar{Y}}, C_R = \frac{S_R}{\bar{R}_X}$$

$$S_X^2 = \frac{\sum_{i=1}^N (X_i - \bar{X})^2}{N-1}, \quad S_Y^2 = \frac{\sum_{i=1}^N (Y_i - \bar{Y})^2}{N-1}$$

$$S_R^2 = \frac{\sum_{i=1}^N (R_{Xi} - \bar{R}_X)^2}{N-1}, \quad S_{XY} = \frac{\sum_{i=1}^N (X_i - \bar{X})(Y_i - \bar{Y})}{N-1}$$

$$S_{XR} = \frac{\sum_{i=1}^N (X_i - \bar{X})(R_{Xi} - \bar{R}_X)}{N-1}, \quad S_{YR} = \frac{\sum_{i=1}^N (Y_i - \bar{Y})(R_{Xi} - \bar{R}_X)}{N-1}$$

จากสมการ (6) เขียนในรูป $e_i : i=1,2,3,4$ ได้ดังสมการ (8)

$$\hat{Y}_{Newd} = k_1 \bar{Y} + k_1 \bar{Y} e_0 + k_2 \bar{X} e_3 - k_2 \bar{X} e_1 + k_3 \bar{R}_X e_4 - k_3 \bar{R}_X e_2 \quad (8)$$

3.1.1. ความเอนเอียง

ความเอนเอียงของตัวประมาณที่นำเสนอเขียน

แทนด้วย $B(\hat{Y}_{Newd})$ ดังสมการ (9)

$$B(\hat{Y}_{Newd}) = (k_1 - 1) \bar{Y} \quad (9)$$

พิสูจน์

$$B(\hat{Y}_{Newd}) = E[\hat{Y}_{Newd} - \bar{Y}]$$

$$= E\left[\left(k_1 \bar{Y} + k_1 \bar{Y} e_0 + k_2 \bar{X} e_3 - k_2 \bar{X} e_1 + k_3 \bar{R}_X e_4 - k_3 \bar{R}_X e_2\right) - \bar{Y}\right]$$

$$= E(k_1 \bar{Y}) + k_1 \bar{Y} E(e_0) + k_2 \bar{X} E(e_3) - k_2 \bar{X} E(e_1) + k_3 \bar{R}_X E(e_4) - k_3 \bar{R}_X E(e_2) - E(\bar{Y})$$

$$= (k_1 - 1) \bar{Y}$$

$$\text{ดังนั้น } B(\hat{Y}_{Newd}) = (k_1 - 1) \bar{Y}$$

เมื่อพิจารณา k_1 เป็นค่าคงที่ใด ๆ แสดงเห็นว่า $\hat{Y}_{Newd}$ เป็นตัวประมาณที่มีความเอนเอียง เพราะมีค่า $B(\hat{Y}_{Newd})$ ไม่เท่ากับ 0 ยกเว้น $k_1 = 1$ จะได้ $\hat{Y}_{Newd}$ เป็นตัวประมาณที่ไม่มีความเอนเอียง เนื่องจาก $B(\hat{Y}_{Newd})$ เท่ากับ 0

3.1.2. ความคลาดเคลื่อนกำลังสองเฉลี่ย

ความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณ

ที่นำเสนอเขียนแทนด้วย $MSE(\hat{Y}_{Newd})$ ดังสมการ (10)

$$MSE(\hat{Y}_{Newd}) = (k_1 - 1)^2 \bar{Y}^2 + k_1^2 \bar{Y}^2 f_1 C_Y^2 + k_2^2 \bar{X}^2 f_3 C_X^2$$

$$+ k_3^2 \bar{R}_X^2 f_3 C_R^2 - 2k_1 k_2 \bar{X} \bar{Y} f_3 \rho_{XY} C_X C_Y$$

$$- 2k_1 k_3 \bar{Y} \bar{R}_X f_3 \rho_{YR} C_Y C_R$$

$$+ 2k_2 k_3 \bar{X} \bar{R}_X f_3 \rho_{XR} C_X C_R \quad (10)$$

พิสูจน์

$$\begin{aligned}
 MSE(\hat{Y}_{Newd}) &= E\left[\left(\hat{Y}_{Newd} - \bar{Y}\right)^2\right] \\
 &= E\left[\left(k_1\bar{Y} + k_1\bar{Y}e_0 + k_2\bar{X}e_3 - k_2\bar{X}e_1 + k_3\bar{R}_X e_4 - k_3\bar{R}_X e_2 - \bar{Y}\right)^2\right] \\
 &= E\left[\left(k_1 - 1\right)^2 \bar{Y}^2 + k_1^2 \bar{Y}^2 e_0^2 + k_2^2 \bar{X}^2 e_3^2 + \dots + 2k_2k_3\bar{X}\bar{R}_X e_1 e_2\right] \\
 &= (k_1 - 1)^2 \bar{Y}^2 + k_1^2 \bar{Y}^2 E[e_0^2] + k_2^2 \bar{X}^2 E[e_3^2] \\
 &\quad + \dots + 2k_2k_3\bar{X}\bar{R}_X E[e_1 e_2]
 \end{aligned}$$

จะได้ $MSE(\hat{Y}_{Newd})$ ดังสมการ (10)

พิจารณาหาค่า k_1, k_2, k_3 ที่ทำให้ $MSE(\hat{Y}_{Newd})$ มีค่าต่ำสุด ดังรายละเอียด

$$\frac{\partial}{\partial k_1} MSE(\hat{Y}_{Newd}) = 0 \quad (11)$$

$$\frac{\partial}{\partial k_2} MSE(\hat{Y}_{Newd}) = 0 \quad (12)$$

$$\frac{\partial}{\partial k_3} MSE(\hat{Y}_{Newd}) = 0 \quad (13)$$

จากสมการ (11)-(13) นำมาแก้ระบบสมการ จะได้

$$k_1 = \frac{W_1}{W_1W_2 + W_3W_4} \quad (14)$$

$$k_2 = \frac{W_5}{W_1W_2 + W_3W_4} \quad (15)$$

$$k_3 = \frac{W_6}{W_1W_2 + W_3W_4} \quad (16)$$

โดยที่

$$W_1 = 1 - \rho_{XR}^2$$

$$W_2 = 1 + f_1C_Y^2$$

$$W_3 = f_3C_Y^2$$

$$W_4 = 2\rho_{YX}\rho_{YR}\rho_{XR} - \rho_{YR}^2 - \rho_{YX}^2$$

$$W_5 = \frac{\bar{Y}C_Y(\rho_{YX} - \rho_{YR}\rho_{XR})}{\bar{X}C_X}$$

$$W_6 = \frac{\bar{Y}C_Y(\rho_{YR} - \rho_{YX}\rho_{XR})}{\bar{R}_XC_R}$$

3.1.3. ศึกษาประสิทธิภาพของตัวประมาณในทางทฤษฎี

การหาประสิทธิภาพของตัวประมาณที่นำเสนอ โดยพิจารณาได้เงื่อนไข MSE ของตัวประมาณที่นำเสนอ มี

ค่าน้อยกว่าตัวประมาณเดิม คือ $\hat{Y}_{Md}$ และ $\hat{Y}_{Rd}$ โดยตัวประมาณ $\hat{Y}_{Md}$ มี $B(\hat{Y}_{Md}) = 0$ ซึ่งเป็นตัวประมาณที่ไม่มี ความเอนเอียง และ Sinsomboonthong [11] ได้แสดงให้เห็นว่า ถ้าตัวประมาณไม่มีความเอนเอียง แล้วความคลาดเคลื่อน กำลังสองเฉลี่ยจะมีค่าเท่ากับ ความแปรปรวน (Variance: Var) ดังนั้น $MSE(\hat{Y}_{Md})$ เท่ากับ $Var(\hat{Y}_{Md})$ คำนวณดัง สมการ (17)

$$Var(\hat{Y}_{Md}) = f_1\bar{Y}^2C_Y^2 \quad (17)$$

ตัวประมาณ $\hat{Y}_{Rd}$ สามารถคำนวณ $MSE(\hat{Y}_{Rd})$ ดังสมการ (18)

$$MSE(\hat{Y}_{Rd}) = \bar{Y}^2 [f_1C_Y^2 + (f_1 - f_2)(C_X^2 - 2\rho_{XY}C_XC_Y)] \quad (18)$$

พบว่า ตัวประมาณ $\hat{Y}_{Newd}$ มีประสิทธิภาพดีกว่าตัวประมาณ $\hat{Y}_{Md}$ ภายใต้เงื่อนไข $MSE(\hat{Y}_{Newd}) < MSE(\hat{Y}_{Md})$ เมื่อ อสมการต่อไปนี้เป็นจริง

$$A_1 > A_2 + A_3 \quad (19)$$

โดยที่

$$A_1 = \bar{Y}^2 [(1 - k_1^2) f_1C_Y^2 - (k_1 - 1)^2]$$

$$A_2 = k_2\bar{X}f_3C_X (k_2\bar{X}C_X + 2k_3\bar{R}_X\rho_{XR}C_R - 2k_1\bar{Y}\rho_{XY}C_Y)$$

$$A_3 = k_3\bar{R}_X f_3C_R (k_3\bar{R}_XC_R - 2k_1\bar{Y}\rho_{YR}C_Y)$$

และตัวประมาณ $\hat{Y}_{Newd}$ มีประสิทธิภาพดีกว่าตัวประมาณ $\hat{Y}_{Rd}$ ภายใต้เงื่อนไข $MSE(\hat{Y}_{Newd}) < MSE(\hat{Y}_{Rd})$ เมื่อ อสมการต่อไปนี้เป็นจริง

$$A_1 + A_4 > A_2 + A_3 \quad (20)$$

โดยที่

$$A_1 = \bar{Y}^2 [(1 - k_1^2) f_1C_Y^2 - (k_1 - 1)^2]$$

$$A_2 = k_2\bar{X}f_3C_X (k_2\bar{X}C_X + 2k_3\bar{R}_X\rho_{XR}C_R - 2k_1\bar{Y}\rho_{XY}C_Y)$$

$$A_3 = k_3\bar{R}_X f_3C_R (k_3\bar{R}_XC_R - 2k_1\bar{Y}\rho_{YR}C_Y)$$

$$A_4 = \bar{Y}^2 f_3 (C_X^2 - 2\rho_{XY}C_XC_Y)$$

3.2. การเปรียบเทียบประสิทธิภาพด้วยข้อมูลเชิงประจักษ์

การศึกษาประสิทธิภาพของตัวประมาณโดยใช้ข้อมูลมลพิษทางอากาศซึ่งพิจารณาจากค่า MSE ของตัวประมาณดังแสดงใน Table 1

จาก Table 1 แสดงผลค่า MSE โดยจำแนกเป็น 3 สถานการณ์ ตามระดับการเลือกตัวอย่างในครั้งแรก คือ 10% 20% และ 50% ตามลำดับ มีรายละเอียดดังนี้

สถานการณ์ที่ 1 เมื่อเลือกตัวอย่างครั้งแรกที่ระดับ 10% จากประชากร หรือ $m = 72$ และเลือกตัวอย่างครั้งที่สองที่ระดับ 20% และ 50% จากการเลือกตัวอย่างในครั้งแรก หรือ $n = 14$ และ 36 ตามลำดับ พบว่า $\hat{Y}_{Newd}$ มีค่า MSE ต่ำสุด

สถานการณ์ที่ 2 เมื่อเลือกตัวอย่างครั้งแรกที่ระดับ 20% จากประชากร หรือ $m = 144$ และเลือกตัวอย่างครั้งที่สองที่ระดับ 20% และ 50% จากการเลือกตัวอย่างในครั้งแรก หรือ $n = 29$ และ 72 ตามลำดับ พบว่า $\hat{Y}_{Newd}$ มีค่า MSE ต่ำสุด

สถานการณ์ที่ 3 เมื่อเลือกตัวอย่างครั้งแรกที่ระดับ 50% จากประชากร หรือ $m = 359$ และเลือกตัวอย่างครั้งที่

สองที่ระดับ 20% และ 50% จากการเลือกตัวอย่างในครั้งแรก หรือ $n = 72$ และ 180 ตามลำดับ พบว่า $\hat{Y}_{Newd}$ มีค่า MSE ต่ำสุด

นอกจากนี้เมื่อพิจารณาว่าค่า MSE จากการศึกษาปัจจัยระดับการเลือกตัวอย่างในแต่ละครั้ง พบว่า

- ระดับการเลือกตัวอย่างครั้งแรกเพิ่มขึ้น ค่า MSE ลดลง ภายใต้ระดับการเลือกตัวอย่างครั้งที่สองเท่ากัน

- ระดับการเลือกตัวอย่างครั้งที่สองเพิ่มขึ้น ค่า MSE ลดลง ภายใต้ระดับการเลือกตัวอย่างครั้งแรกเท่ากัน

จากผลการทดลองในทุกสถานการณ์สอดคล้องกัน โดยแสดงให้เห็นว่าการใช้ลำดับของตัวแปรเสริมสามารถปรับปรุงประสิทธิภาพของตัวประมาณได้เนื่องจาก $\hat{Y}_{Newd}$ มีค่า MSE ต่ำสุด ซึ่งสอดคล้องกับงานวิจัยของ Ullah et al. [5], Ajayi et al. [6], Tahir et al. [7], Çetin and Koyuncu [8] และ Haq et al. [9] นอกจากนี้ระดับการสุ่มตัวอย่างหรือขนาดตัวอย่างมีผลต่อประสิทธิภาพของตัวประมาณ เมื่อระดับการเลือกตัวอย่างแต่ละครั้งเพิ่มขึ้น ทุกตัวประมาณมีค่า MSE ลดลง ซึ่งสอดคล้องกับงานวิจัยของ Nathomthong and Chutiman [12]

Table 1 Mean square error (MSE) of each estimator under different situations

Level of sample size		MSE		
% (m)	% (n)	$\hat{Y}_{Md}$	$\hat{Y}_{Rd}$	$\hat{Y}_{Newd}$
10 (72)	20 (14)	31.9644	30.9572	12.8426
	50 (36)	16.3051	15.3026	8.7396
20 (144)	20 (29)	18.6841	17.6800	6.2474
	50 (72)	6.2970	5.2961	3.5543
50 (359)	20 (72)	7.4705	6.4688	2.0717
	50 (180)	1.9273	1.7270	0.8614

4. บทสรุป

งานวิจัยนี้นำเสนอการพัฒนาตัวประมาณค่าเฉลี่ยประชากรโดยใช้ลำดับของตัวแปรเสริมภายใต้การเลือกตัวอย่างสองครั้ง ผลการศึกษาทางทฤษฎีและการใช้ข้อมูลเชิงประจักษ์มีความสอดคล้องกัน โดยแสดงให้เห็นว่าตัวประมาณที่นำเสนอมีประสิทธิภาพดีกว่าตัวประมาณเดิมภายใต้เงื่อนไขที่ได้จากทฤษฎี ดังนั้นการใช้ข้อมูลลำดับของตัวแปรเสริมสามารถเพิ่มประสิทธิภาพของตัวประมาณค่าเฉลี่ยได้นอกจากนี้ขนาดตัวอย่างที่เพิ่มขึ้นมีผลต่อประสิทธิภาพของ

ตัวประมาณ เมื่อขนาดตัวอย่างใหญ่ขึ้นประสิทธิภาพของตัวประมาณเพิ่มขึ้น ผลลัพธ์ของการศึกษานี้สามารถนำไปใช้กับงานวิจัยที่มีการสุ่มตัวอย่างแบบสองครั้งและมีการตรวจสอบเงื่อนไขจากทฤษฎีเป็นจริงจึงสามารถใช้ตัวประมาณที่นำเสนอได้อย่างเหมาะสม

ข้อเสนอแนะของการศึกษาค้างต่อไปอาจมีการพัฒนาตัวประมาณด้วยการใช้ตัวแปรเสริม 2 ตัวแปร ภายใต้กรณีไม่ทราบค่าประชากรด้วยการประยุกต์ใช้ลำดับของตัวแปรเสริมควบคู่ในการพัฒนาตัวประมาณ นอกจากนี้อาจใช้ข้อมูล

คุณลักษณะของตัวแปรเสริม (Auxiliary attribute) ในการพัฒนาตัวประมาณภายใต้การเลือกตัวอย่างสองครั้ง หรือศึกษาภายใต้รูปแบบการสุ่มตัวอย่างอื่น ๆ เช่น การสุ่มตัวอย่างแบบแบ่งชั้น (Stratified sampling) การสุ่มตัวอย่างแบบกลุ่ม (Cluster sampling)

5. References

- [1] Khan, M. 2016. A ratio chain-type exponential estimator for finite population mean using double sampling. **SpringerPlus**. 5: 86.
- [2] Sharma, V. and Kumar, S. 2021. Class of ratio-cum-product type estimator under double sampling: a simulation study. **Thailand Statistician**. 19(4): 734-742.
- [3] Vadlamudi, K.A. Sedory, S.A. and Singh, S. 2017. A new estimator of mean using double sampling. **Statistics in Transition New Series**. 18(4): 271-290.
- [4] Khare, B.B. and Rehman, H.U. 2013. Improved chain type estimators for population mean using two auxiliary variables and double sampling scheme. **International Journal of Statistics and Probability**. 1(3): 82-87.
- [5] Ullah, K. and et al. 2022. Estimation of finite population mean in simple and stratified random sampling by utilizing the auxiliary, ranks, and square of the auxiliary information. **Mathematical Problems in Engineering**. 2022: 263492.
- [6] Ajayi, A.O. and et al. 2023. A modified generalized class of exponential ratio type estimators in ranked set sampling. **Scientific African**. 19: e01447.
- [7] Tahir, M. and et al. 2023. A new improved estimator for the population mean using twofold auxiliary information under simple random sampling. **Management Science Letters**. 13(4): 265-276.
- [8] Cetin, A.E. and Koyuncu, N. 2024. New robust class of estimators for population mean under different sampling designs. **Journal of Computational and Applied Mathematics**. 441: 115669.
- [9] Haq, A., Khan, M. and Hussian, Z. 2017. A new estimator of finite population mean based on the dual use of the auxiliary information. **Communications in Statistics - Theory and Methods**. 46(9): 4425-4436.
- [10] Pollution Control Department. 2019. **Historical Data of Thailand's Air Quality**. <http://air4thai.pcd.go.th/webV3/#/History>. Accessed 5 February 2019. (in Thai)
- [11] Sinsomboonthong, S. 2011. **Mathematical Statistics 1**. Bangkok: Chulalongkorn University Press. (in Thai)
- [12] Nathomthong, A. and Chutiman, N. 2024. Unbiased estimators using auxiliary information for the finite population mean under two-phase sampling. **Lobachevskii Journal of Mathematics**. 45: 741-749.