ตัวประมาณอัตราส่วนร่วมกับผลคูณแบบปรับปรุงภายใต้การเลือกตัวอย่างสองเฟส โดยใช้ตัวแปรช่วยสองตัวแปร

Modified Ratio-cum-product Estimators under Two-phase Sampling using Two Auxiliary Variables

จิราภรณ์ ธรรมสาโรช, ภาณุวัฒน์ พิมพ์ทรัพย์ และวุฒิชัย ศรีโสดาพล*

ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น อำเภอเมือง จังหวัดขอนแก่น 40002

*Email: wuttsr@kku.ac.th

บทคัดย่อ

วัตถุประสงค์ของการศึกษาคือเสนอตัวประมาณอัตราส่วนร่วมกับผลคูณสำหรับการประมาณค่าเฉลี่ยของประชากรภายใต้การเลือกตัวอย่างสองเฟสโดยใช้ตัวแปรช่วยสองตัวแปร ตัวประมาณที่นำเสนอปรับปรุงตัวประมาณของ Singh และคณะ ในปี 2005 และตัวประมาณของ Chanu และ Singh ในปี 2014 พร้อมทั้งหาความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณที่นำเสนอ ทำการเปรียบเทียบประสิทธิภาพของตัวประมาณที่นำเสนอกับตัวประมาณที่ไม่เอนเอียง ตัวประมาณของ Singh และคณะ ในปี 2005 และตัวประมาณของ Chanu และ Singh ในปี 2014 โดยใช้ความคลาดเคลื่อนกำลังสองเฉลี่ย ผลการศึกษาพบว่าตัวประมาณที่นำเสนอมีประสิทธิภาพมากกว่าตัวประมาณที่ไม่เอนเอียง ตัวประมาณของ Singh และคณะ ในปี 2005 และตัวประมาณของ Chanu และ Singh ในปี 2014 ภายใต้การศึกษาจากข้อมูลจริง

คำสำคัญ : ตัวแปรช่วย ตัวประมาณอัตราส่วนร่วมกับผลคูณ ประสิทธิภาพ ความคลาดเคลื่อนกำลังสองเฉลี่ย การเลือกตัวอย่างสองเฟส

Abstract

The objective of this study was to propose the ratio-cum-product estimators of a population mean under two-phase sampling using two auxiliary variables. These proposed estimators were modified from the estimators of Singh et al. (2005) and Chanu and Singh (2014). The mean squared errors of the proposed estimators were obtained. The efficiencies of these proposed estimators were compared with a usual unbiased estimator and the estimators of Singh et al. (2005) and Chanu and Singh (2014) based on mean squared error. The results showed that the proposed estimators were more efficient than the usual unbiased estimator and the estimators of Singh et al. (2005) and Chanu and Singh (2014) under empirical study with some real dataset.

Keywords: Auxiliary; variable; Ratio-cum-product estimator; Efficiency; Mean squared error; Two-phase sampling

บทนำ

ในการวิจัยเชิงสำรวจ ผู้วิจัยต้องทำการเก็บรวบรวมข้อมูลของประชากรที่ต้องการทำการวิจัยนั้นอาจมีเพียงกลุ่มเดียวหรือหลายกลุ่มขึ้นอยู่กับวัตถุประสงค์ในการวิจัย ซึ่งในกรณีที่ไม่อาจศึกษาประชากรได้ทั้งหมดจะต้องทำการเลือกตัวอย่างเพื่อหาตัวแทนที่ดีจำนวนหนึ่งของประชากรมาใช้ในการศึกษาโดยตัวอย่างนั้นจะต้องให้ข้อมูลและสารสนเทศของประชากรที่จะทำการศึกษาได้อย่างครบถ้วนโดยวิธีที่เหมาะสมที่จะทำให้ได้ข้อมูลที่ใกล้เคียงกับประชากรมากที่สุดคือ การเลือกตัวอย่างที่เป็นตัวแทนที่ดีของประชากร ซึ่งต้องอาศัยวิธีการทางสถิติที่เรียกว่าเทคนิคการเลือกตัวอย่าง (Sampling Technique) เข้ามาช่วยเพื่อให้ได้ตัวอย่างที่มีความใกล้เคียงกับประชากรมากที่สุด

การเลือกตัวอย่างเชิงสำรวจจะทำการอนุมานค่าของประชากรโดยการนำเอาสารสนเทศมาจากตัวอย่างที่เลือกจากประชากรมาใช้ในการวิเคราะห์และอธิบายถึงคุณลักษณะต่างๆของประชากรซึ่งเรียกว่าพารามิเตอร์ (Parameter) ซึ่งคุณลักษณะของประชากรเหล่านั้นจะอยู่ในรูปแบบต่างๆ เช่น ผลรวมประชากร ค่าเฉลี่ยประชากร สัดส่วนประชากร และความแปรปรวนของประชากร เป็นต้น ซึ่งค่าเฉลี่ยประชากรจะเป็นพารามิเตอร์ที่ถูกสนใจศึกษาหรือถูกประมาณค่า โดยทำการเก็บรวบรวมข้อมูลจากตัวอย่างซึ่งเป็นส่วนหนึ่งของประชากรด้วยวิธีเลือกตัวอย่างแล้วนำข้อมูลนั้นไปทำการประมาณค่าด้วยตัวประมาณ (Estimator) ซึ่งเป็นปัจจัยสำคัญที่ใช้ในการวิเคราะห์และอธิบายถึงคุณลักษณะต่างๆของประชากรที่สนใจทำการศึกษา

เมื่อพิจารณาการเลือกตัวอย่างสุ่มแบบง่ายโดยทำการเลือกตัวอย่างครั้งเดียวด้วยวิธีการเลือกแบบไม่ใส่คืนขนาด n จากประชากรขนาด N โดยที่ค่าเฉลี่ยประชากร $\bar{Y} = \frac{\sum_{i=1}^{N} Y_i}{N}$ จะถูกประมาณด้วยค่าเฉลี่ยตัวอย่าง $\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$ ซึ่งเป็นตัวประมาณที่ไม่เอนเอียง (Unbiased Estimator) ของ $\bar{Y}$ โดยความแปรปรวน คือ

$$Var(\bar{y}) = \frac{1-f}{n}\bar{Y}^2 C_y^2$$

โดยที่ $f = \frac{n}{N}$ คือเศษส่วนการเลือกตัวอย่าง (Sampling Fraction)

C_y คือสัมประสิทธิ์การแปรผัน (Coefficient of Variation) ของตัวแปรที่ทำการศึกษา y

นอกจากตัวประมาณของประชากรดังกล่าวแล้ว ยังมีตัวประมาณอีกรูปแบบหนึ่งซึ่งนำเอาสารสนเทศของตัวแปรช่วย x มาช่วยเพิ่มประสิทธิภาพในการประมาณค่าเฉลี่ยของประชากรในตัวแปรที่ทำการศึกษา y โดยตัวแปรช่วย x ต้องมีความสัมพันธ์ทางบวกกับตัวแปรที่ต้องการศึกษา y ในระดับสูงและเรียกตัวประมาณดังกล่าวว่า ตัวประมาณแบบอัตราส่วน (Ratio Estimator) ซึ่งมีรูปแบบดังนี้

$$\bar{y}_R = \bar{y}\left(\frac{\bar{X}}{\bar{x}}\right)$$

เมื่อ $\bar{x}$ คือค่าเฉลี่ยตัวอย่างของตัวแปรช่วย x

$\bar{y}$ คือค่าเฉลี่ยตัวอย่างของตัวแปร y

โดยมีความคลาดเคลื่อนกำลังสองเฉลี่ย คือ

$$MSE(\bar{y}_R) = \left(\frac{1-f}{n}\bar{Y}^2\right)\left[C_y^2 + C_x^2\left(1-2K_{yx}\right)\right]$$

เมื่อ C_y คือสัมประสิทธิ์การแปรผัน (Coefficient of Variation) ของตัวแปรที่ทำการศึกษา y

C_x คือสัมประสิทธิ์การแปรผัน (Coefficient of Variation) ของตัวแปรช่วย x

และ $K_{yx} = \frac{\rho_{yx} C_y}{C_x}$

ในทางกลับกัน ถ้าตัวแปรช่วย x มีความสัมพันธ์ทางลบกับตัวแปรที่ต้องการศึกษา y ในระดับสูงจะเรียกตัวประมาณนี้ว่า ตัวประมาณผลคูณ (Product Estimator) ซึ่งมีรูปแบบดังนี้

$$\bar{y}_P = \bar{y}\left(\frac{\bar{x}}{\bar{X}}\right)$$

โดยมีความคลาดเคลื่อนกำลังสองเฉลี่ย คือ

$$MSE(\bar{y}_P) = \left(\frac{1-f}{n}\bar{Y}^2\right)\left[C_y^2 + C_x^2\left(1-2K_{yx}\right)\right]$$

นอกจากตัวประมาณแบบอัตราส่วนและตัวประมาณแบบผลคูณดังกล่าวข้างต้นแล้วยังมีการนำเสนอตัวประมาณแบบอื่นในการประมาณค่าเฉลี่ยประชากร เช่น

ในปี 1967 Singh [1] ได้นำเสนอตัวประมาณแบบอัตราส่วนร่วมกับผลคูณ (Ratio Cum Product Estimator) ซึ่งเป็นการนำตัวประมาณแบบอัตราส่วนมาผสมกับตัวประมาณแบบผลคูณเพื่อเพิ่มประสิทธิภาพของการประมาณค่าโดยใช้ตัวแปรช่วยสองตัว คือ ตัวแปรช่วย x ซึ่งมีความสัมพันธ์ทางบวกกับตัวแปรที่สนใจศึกษา y และตัวแปรช่วย z ซึ่งมีความสัมพันธ์ทางลบกับตัวแปรที่สนใจศึกษา y มาใช้ในการสร้างตัวประมาณ ตัวประมาณที่นำเสนอมีรูปแบบดังนี้

$$\bar{y}_{RP} = \bar{y}\left(\frac{\bar{X}}{\bar{x}}\frac{\bar{z}}{\bar{Z}}\right)$$

เมื่อ $\bar{Z}$ คือค่าเฉลี่ยประชากรของตัวแปรช่วย z

$\bar{z}$ คือค่าเฉลี่ยตัวอย่างของตัวแปรช่วย z

โดยมีความคลาดเคลื่อนกำลังสองเฉลี่ย คือ

$$MSE(\bar{y}_{PR}) = \left(\frac{1-f}{n}\bar{Y}^2\right)\left[C_y^2 + C_z^2\left(1-2K_{yz}\right) + C_x^2\left(1-2K\right)\right]$$

เมื่อ C_z คือสัมประสิทธิ์การแปรผัน (Coefficient of Variation) ของตัวแปรช่วย z

และ $K_{yz} = \dfrac{\rho_{yz}C_y}{C_z}$, $K_{zx} = \dfrac{\rho_{xz}C_z}{C_x}$, $K = K_{yx} + K_{zx}$

ต่อมาในปี 1996 Tracy และคณะ [2] ได้ทำการปรับปรุงตัวประมาณของ Singh ในปี 1967 [1] โดยทำการแปลงให้อยู่ในรูปของ $u_i = L - x_i$ เมื่อ L คือค่าคงที่ที่เหมาะสมสำหรับการประมาณค่าของตัวประมาณที่นำเสนอ

ในปี 2005 Singh และคณะ [3] ได้ปรับปรุงตัวประมาณของ Singh ในปี 1967 [1] โดยทำการแปลงค่าให้ x_i และ z_i อยู่ในรูปแบบของ $x_i^* = (1+g)\bar{X} - gx_i$ และ $z_i^* = (1+g)\bar{Z} - gz_i$ เมื่อ $g = n/(N-n)$ เพื่อปรับปรุงตัวประมาณที่นำเสนอให้มีประสิทธิภาพมากยิ่งขึ้น ซึ่งตัวประมาณที่นำเสนอมีรูปแบบดังนี้

$$\bar{y}_{RP}^* = \bar{y}\left(\frac{\bar{x}^*}{\bar{X}}\frac{\bar{Z}}{\bar{z}^*}\right)$$

โดยมีความคลาดเคลื่อนกำลังสองเฉลี่ย คือ

$$MSE(\bar{y}_{RP}^*) = \frac{1-f}{n}\bar{Y}^2\left[C_y^2 + gC_z^2\left(g+2K_{yz}\right) + gC_x^2\left(g-2gK_{zx}-2K_{yx}\right)\right]$$

ในปีเดียวกัน Singh และ Tailor [4] ได้ทำการพัฒนาตัวประมาณของ Singh ในปี 1967 [1] โดยการนำสัมประสิทธิ์สหสัมพันธ์ (ρ_{xz}) ของตัวแปรช่วย x และ z มาใช้เพื่อเพิ่มประสิทธิภาพของตัวประมาณที่นำเสนอ

จากนั้นในปี 2010 Tailor และคณะ [5] ได้นำเอาค่าสัมประสิทธิ์ความโด่ง ($\beta_2(x)$) ของตัวแปรช่วย x และ $\beta_2(z)$ ของตัวแปรช่วย z เข้ามาปรับปรุงตัวประมาณของ Singh ในปี 1967 [1] เพื่อเพิ่มประสิทธิภาพในการประมาณค่า ซึ่งต่อมาในปี 2011 Tailor และคณะ [6] ได้มีการปรับปรุงตัวประมาณของ Singh ในปี 1967 [1] เช่นกันโดยนำเอาสัมประสิทธิ์การแปรผัน (C_x) ของตัวแปรช่วย x สัมประสิทธิ์การแปรผัน (C_z) ของตัวแปรช่วย z สัมประสิทธิ์ความโด่งของตัวแปรช่วย x และสัมประสิทธิ์ความโด่งของตัวแปรช่วย z มาใช้ร่วมกันในการเพิ่มประสิทธิภาพของตัวประมาณที่นำเสนอ และในปี 2012 Tailor และคณะ [7] ได้นำเอาสัมประสิทธิ์สหสัมพันธ์ของตัวแปรช่วย x และ z (ρ_{xz}) เข้ามาใช้ร่วมกับการแปลงของ x_i และ z_i ที่อยู่ในรูปแบบของ $x_i^* = (1+g)\bar{X} - gx_i$ และ $z_i^* = (1+g)\bar{Z} - gz_i$ เพื่อทำการปรับปรุงตัวประมาณซึ่งทำให้ตัวประมาณมีความเหมาะสมและมีประสิทธิภาพมากยิ่งขึ้น

เมื่อตัวแปรช่วย x และ z ไม่สามารถหาสารสนเทศได้ในระดับประชากร เช่น ผลรวมประชากร ค่าเฉลี่ยประชากร และสัดส่วนประชากร เป็นต้น จึงต้องทำการเลือกตัวอย่างสองเฟส (Two-phase sampling) เพื่อใช้คำนวณหาสารสนเทศดังกล่าว โดยทำการเลือกตัวอย่างเฟสที่ 1 ขนาด n_1 จากประชากรขนาด N แล้วทำการการเลือกหน่วยตัวอย่างเฟสที่ 2 ขนาด n จากประชากรขนาด N อีกครั้งโดยตัวอย่างจากเฟสที่ 2 ไม่ได้เป็นตัวอย่างย่อยในเฟสที่ 1 เมื่อ $n_1 < N$, $n < N$ และ $n < n_1$ จะทำให้หน่วยตัวอย่างสำหรับการประมาณค่าเฉลี่ยประชากรอิสระ

กัน และถ้าหากเลือกตัวอย่างเฟสที่ 2 ขนาด n จากตัวอย่างขนาด n_1 โดยตัวอย่างจากเฟสที่ 2 เป็นตัวอย่างย่อยในเฟสที่ 1 เมื่อ $n_1 < N$ และ $n < n_1$ ซึ่งจะทำให้หน่วยตัวอย่างมีความสัมพันธ์กัน ทำให้ในปี 2014 Chanu และ Singh [8] ได้ทำการพัฒนาตัวประมาณของ Singh ในปี 1967 [1] โดยนำเสนอตัวประมาณในรูปของการเลือกตัวอย่างสองเฟสแล้วให้ตัวประมาณที่นำเสนอปรับให้อยู่ในรูปของการยกกำลังด้วย α แล้วหาค่า α ซึ่งเป็นค่าคงที่ที่เหมาะสมที่สุดในการประมาณค่า ในกรณีหน่วยตัวอย่างอิสระกันตัวประมาณที่นำเสนอมีรูปแบบดังนี้

$$\bar{Y}_{RP}^{\omega(d)} = \bar{y}\left(\frac{\bar{x}_1}{\bar{x}}\frac{\bar{z}}{\bar{z}_1}\right)^{\alpha}$$

โดยมีความคลาดเคลื่อนกำลังสองเฉลี่ย คือ

$$MSE\left(\bar{Y}_{RP}^{\omega(d)}\right) = \bar{Y}^2 f^{'} C_y^2 + \bar{Y}^2 2\left[\alpha^2 N_1 K_3 - \alpha f^{'} S_1\right]$$

สำหรับกรณีหน่วยตัวอย่างอิสระกันตัว ประมาณที่นำเสนอมีรูปแบบดังนี้

$$\bar{Y}_{RP}^{\omega d} = \bar{y}\left(\frac{\bar{x}_1}{\bar{x}}\frac{\bar{z}}{\bar{z}_1}\right)^{\alpha}$$

โดยมีความคลาดเคลื่อนกำลังสองเฉลี่ย คือ

$$MSE\left(\bar{Y}_{RP}^{\omega d}\right) = \bar{Y}^2\left(\frac{1-f}{n}\right)C_y^2 + \bar{Y}^2\left(\frac{1-f_1}{n}\right)\left[\alpha^2 K_3 - 2\alpha S_1\right]$$

เมื่อ $f_1 = \frac{n}{n_1}$, $S_1 = C_{yx-}C_{yz}$,

$K_3 = C_x^2 + C_z^2(1-2K_{xz})$

และ $N_1 = \frac{1}{2}\left(\frac{1}{n} + \frac{1}{n_1} - \frac{2}{N}\right)$

จากการศึกษาข้างต้นผู้วิจัยจึงทำการปรับปรุงตัวประมาณอัตราส่วนร่วมกับผลคูณ ของ Singh และคณะ [3] ในปี 2005 และตัวประมาณอัตราส่วนร่วมกับผลคูณของ Chanu และ Singh ในปี 2014 [8] ให้มีประสิทธิภาพมากยิ่งขึ้น โดยทำการปรับปรุงตัวประมาณภายใต้การเลือกตัวอย่างสุ่มแบบง่ายสองเฟสแบบไม่ใส่คืน ซึ่งจะนำเอาสัมประสิทธิ์การแปรผันของตัวแปรช่วย x สัมประสิทธิ์การแปรผันของตัวแปรช่วย z สัมประสิทธิ์ความโด่งของตัวแปรช่วย x สัมประสิทธิ์ความโด่งของตัวแปรช่วย z และสัมประสิทธิ์สหสัมพันธ์ของตัวแปรช่วย x และ z มาใช้ในการปรับปรุงตัวประมาณ โดยผู้วิจัยได้ทำการแบ่งตัวประมาณในการนำเสนอออกเป็น 2 กรณีคือ กรณีหน่วยตัวอย่างไม่เป็นอิสระกัน โดยตัวอย่างจากเฟสที่ 2 เป็นตัวอย่างย่อยในเฟสที่ 1 และกรณีหน่วยตัวอย่างเป็นอิสระกัน โดยตัวอย่างจากเฟสที่ 2 ไม่ได้เป็นตัวอย่างย่อยในเฟสที่ 1 เมื่อทำการเลือกหน่วยตัวอย่างขนาด n_1 และ n โดยที่ $n_1 < N$ และ $n < n_1$

วิธีดำเนินการวิจัย

การปรับปรุงตัวประมาณสำหรับการประมาณค่าเฉลี่ยประชากร โดยทำการปรับปรุงตัวประมาณอัตราส่วนร่วมกับผลคูณของ Singh ในปี 1967 [1] ตัวประมาณของ Singh และคณะในปี 2005 [3] และตัวประมาณของ Chanu และ Singh ในปี 2014 [8] แล้วทำการเปรียบเทียบประสิทธิภาพของตัวประมาณที่ทำการปรับปรุงจากความคลาดเคลื่อนกำลังสองเฉลี่ยโดยมีวิธีดำเนินการดังต่อไปนี้

การนำเสนอตัวประมาณ

1. ทำการปรับปรุงตัวประมาณอัตราส่วนร่วมกับผลคูณของ Singh และคณะในปี 2005 [3] โดยผู้วิจัยได้ทำการแปลงค่าของ x_i และ $z_i \; ; i = 1,2,...,N$ ให้อยู่ในรูปของ $x_i^* = (1+g)\bar{x}_1 - gx_i$ และ $z_i^* = (1+g)\bar{z}_1 - gz_i$ มาใช้ร่วมกับสัมประสิทธิ์การแปรผันของตัวแปรช่วย x สัมประสิทธิ์การแปรผันของตัวแปรช่วย z สัมประสิทธิ์ความโด่งของตัวแปรช่วย x สัมประสิทธิ์ความโด่งของตัวแปรช่วย z และสัมประสิทธิ์สหสัมพันธ์ของตัวแปรช่วย x และ z ในการปรับปรุงตัวประมาณ ทำให้ได้ตัวประมาณอัตราส่วนร่วมกับผลคูณในการประมาณค่าเฉลี่ยประชากร คือ

$$\bar{y}_{1d} = \bar{y}\left[\left(\frac{\bar{x}^* + C_x}{\bar{x}_1 + C_x}\right)\left(\frac{\bar{z}_1 + C_z}{\bar{z}^* + C_z}\right)\right] \quad (1)$$

$$\bar{y}_{2d} = \bar{y}\left[\left(\frac{\bar{x}^* + \beta_2(x)}{\bar{x}_1 + \beta_2(x)}\right)\left(\frac{\bar{z}_1 + \beta_2(z)}{\bar{z}^* + \beta_2(z)}\right)\right] \quad (2)$$

และ

$$\bar{y}_{3d} = \bar{y}\left[\left(\frac{\bar{x}^* + \rho_{xz}}{\bar{x}_1 + \rho_{xz}}\right)\left(\frac{\bar{z}_1 + \rho_{xz}}{\bar{z}^* + \rho_{xz}}\right)\right] \quad (3)$$

เมื่อ $\bar{y}$ คือค่าเฉลี่ยตัวอย่างของตัวแปรที่สนใจศึกษา y

$\bar{x}_1$ คือค่าเฉลี่ยตัวอย่างของตัวแปรช่วย x จากการเลือกหน่วยตัวอย่างในเฟสที่ 1

$\bar{x}$ คือค่าเฉลี่ยตัวอย่างของตัวแปรช่วย x จากการเลือกหน่วยตัวอย่างในเฟสที่ 2

$\bar{z}_1$ คือค่าเฉลี่ยตัวอย่างของตัวแปรช่วย z จากการเลือกหน่วยตัวอย่างในเฟสที่ 1

$\bar{z}$ คือค่าเฉลี่ยตัวอย่างของตัวแปรช่วย z จากการเลือกหน่วยตัวอย่างในเฟสที่ 2

C_x คือสัมประสิทธิ์การแปรผันประชากรของตัวแปรช่วย x

C_z คือสัมประสิทธิ์การแปรผันประชากรของตัวแปรช่วย z

$\beta_2(x)$ คือสัมประสิทธิ์ความโด่งประชากรของตัวแปรช่วย x

$\beta_2(z)$ คือสัมประสิทธิ์ความโด่งประชากรของตัวแปรช่วย z

ρ_{xz} คือสัมประสิทธิ์สหสัมพันธ์ประชากรของตัวแปรช่วย x และ z

2. ทำการปรับปรุงตัวประมาณอัตราส่วนร่วมกับผลคูณของ Chanu และ Singh ในปี 2014 [8] โดยนำสัมประสิทธิ์การแปรผันของตัวแปรช่วย x สัมประสิทธิ์การแปรผันของตัวแปรช่วย z สัมประสิทธิ์ความโด่งของตัวแปรช่วย x สัมประสิทธิ์ความโด่งของตัวแปรช่วย z และสัมประสิทธิ์สหสัมพันธ์ของตัวแปรช่วย x และ z มาใช้ในการปรับปรุงตัวประมาณ จากนั้นทำให้ตัวประมาณอยู่ในรูปของการยกกำลังด้วย α ซึ่ง α คือค่าคงที่ที่เหมาะสมในการประมาณค่า ทำให้ได้ตัวประมาณอัตราส่วนร่วมกับผลคูณสำหรับประมาณค่าเฉลี่ยประชากร คือ

$$\bar{y}_{4d} = \bar{y}\left(\left(\frac{\bar{x}_1 + C_x}{\bar{x} + C_x}\right)\left(\frac{\bar{z} + C_z}{\bar{z}_1 + C_z}\right)\right)^{\alpha} \quad (4)$$

$$\bar{y}_{5d} = \bar{y}\left(\left(\frac{\bar{x}_1 + \beta_2(x)}{\bar{x} + \beta_2(x)}\right)\left(\frac{\bar{z} + \beta_2(z)}{\bar{z}_1 + \beta_2(z)}\right)\right)^{\alpha} \quad (5)$$

และ

$$\bar{y}_{6d} = \bar{y}\left(\left(\frac{\bar{x}_1 + \rho_{xz}}{\bar{x} + \rho_{xz}}\right)\left(\frac{\bar{z} + \rho_{xz}}{\bar{z}_1 + \rho_{xz}}\right)\right)^{\alpha} \quad (6)$$

เมื่อ α เป็นค่าคงที่ที่เหมาะสมสำหรับประมาณค่า

ความเอนเอียงและความความเคลื่อนกำลังสองเฉลี่ยของตัวประมาณ

ในการหาความเอนเอียง (Bias) ของตัวประมาณ $Bias(\bar{y}_{id}) = E(\bar{y}_{id} - \bar{Y})$ และความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) ของตัวประมาณ $MSE(\bar{y}_{id}) = E(\bar{y}_{id} - \bar{Y})^2$; $i = 1, 2, ..., 6$ จะทำการแปลงให้ตัวประมาณที่นำเสนออยู่ในรูปของ e [8] นั่นคือ

$\bar{y} = \bar{Y}(1+e_0)$, $\bar{x} = \bar{X}(1+e_1)$, $\bar{x}_1 = \bar{X}(1+e_2)$, $\bar{z} = \bar{Z}(1+e_3)$ และ $\bar{z}_1 = \bar{Z}(1+e_4)$

โดยที่

$e_0 = (\bar{y} - \bar{Y})/\bar{Y}$, $e_1 = (\bar{x} - \bar{X})/\bar{X}$,

$e_2 = (\bar{x}_1 - \bar{X})/\bar{X}$ $e_3 = (\bar{z} - \bar{Z})/\bar{Z}$ และ

$e_4 = (\bar{z}_1 - \bar{Z})/\bar{Z}$

ซึ่งการหาความเอนเอียงและความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณที่นำเสนอมีเงื่อนไขแบ่งออกเป็น 2 กรณี ดังต่อไปนี้

กรณีที่ 1 เมื่อทำการเลือกตัวอย่างเฟสที่ 2 ขนาด n จากตัวอย่างเฟสที่ 1 ขนาด n_1 โดยที่ $n < n_1$ โดยตัวอย่างจากเฟสที่ 2 เป็นตัวอย่างย่อยในเฟสที่ 1 เมื่อ

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

$$E(e_0^2) = \frac{1-f}{n}C_y^2,\ E(e_1^2) = \frac{1-f}{n}C_x^2,$$

$$E(e_3^2)=\frac{1-f}{n}C_z^2\,,$$

$$E(e_2^2)=E(e_1e_2)=\frac{1-f^{'}}{n}C_x^2\,,$$

$$E(e_4^2)=E(e_3e_4)=\frac{1-f^{'}}{n}C_z^2\,,$$

$$E(e_0e_1)=\frac{1-f}{n}\rho_{yx}C_yC_x\,,$$

$$E(e_0e_2)=\frac{1-f^{'}}{n}\rho_{yx}C_yC_x\,,$$

$$E(e_0e_3)=\frac{1-f}{n}\rho_{yz}C_yC_z\,,$$

$$E(e_0e_4)=\frac{1-f^{'}}{n}\rho_{yz}C_yC_z$$

$$E(e_1e_3)=\frac{1-f}{n}\rho_{xz}C_xC_z\,,$$

$$E(e_1e_4)=E(e_2e_3)=E(e_2e_4)=\frac{1-f^{'}}{n}\rho_{xz}C_xC_z$$

เมื่อ $f=n/N$ และ $f^{'}=n_1/N$

กรณีที่ 2 เมื่อทำการเลือกตัวอย่างเฟสที่ 2 ขนาด n จากตัวอย่างขนาด N โดยที่ $n<n_1$ โดยตัวอย่างจากเฟสที่ 2 ไม่ได้เป็นตัวอย่างย่อยในเฟสที่ 1 เมื่อ

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

$$E(e_0^2)=\frac{1-f}{n}C_y^2\,,\ E(e_1^2)=\frac{1-f}{n}C_x^2\,,$$

$$E(e_3^2)=\frac{1-f}{n}C_z^2\,,$$

$$E(e_2^2)=E(e_1e_2)=\frac{1-f^{'}}{n}C_x^2\,,$$

$$E(e_4^2)=E(e_3e_4)=\frac{1-f^{'}}{n}C_z^2\,,$$

$$E(e_0e_1)=\frac{1-f}{n}\rho_{yx}C_yC_x\,,$$

$$E(e_0e_3)=\frac{1-f}{n}\rho_{yz}C_yC_z\,,$$

$$E(e_1e_3)=\frac{1-f}{n}\rho_{xz}C_xC_z\,,$$

$$E(e_2e_4)=\frac{1-f^{'}}{n}\rho_{xz}C_xC_z$$

$$E(e_0e_2)=E(e_0e_4)=E(e_1e_2)=0$$

$$E(e_1e_4)=E(e_2e_3)=E(e_3e_4)=0$$

ผลการวิจัย

1. ความเอนเอียงและความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณที่นำเสนอ

นำตัวประมาณที่นำเสนอ $\bar{y}_{id}$ $;i=1,2,...,6$ ที่ในรูปของ e ไปหาความเอนเอียงและความคลาดเคลื่อนกำลังสองเฉลี่ย จาก $Bias(\bar{y}_{id})=E(\bar{y}_{id}-\bar{Y})$ และ $MSE(\bar{y}_{id})=E(\bar{y}_{id}-\bar{Y})^2$ ตามลำดับ แบ่งตามกรณีที่ 1 และกรณีที่ 2 ทำให้ได้ความเอนเอียงและความคลาดเคลื่อนกำลังสองเฉลี่ยดังต่อไปนี้

ความเอนเอียงของตัวประมาณที่นำเสนอในกรณีที่ 1 คือ

$$Bias(\bar{y}_{1d})=\bar{Y}\left(\frac{1-f}{n}\right)C_z^{\,2}\left(-\theta_1\theta_2g^2K_{xz}+\theta_2^{\,2}g+\theta_2gK_{yz}\right)+\bar{Y}\left(\frac{1-f'}{n}\right)\Big[C_x^{\,2}\left(-2\theta_1^2g+\theta_1gK_{yx}\right)$$
$$+C_z^2\left(-\theta_1\theta_2g^2K_{xz}-\theta_2^2g^2+2\theta_1\theta_2g^2K_{xz}-\theta_2gK_{yz}\right)\Big]$$

$$Bias(\bar{y}_{2d})=\bar{Y}\left(\frac{1-f}{n}\right)C_z^{\,2}\left(-t_1t_2g^2K_{xz}+t_2^{\,2}g+t_2gK_{yz}\right)+\bar{Y}\left(\frac{1-f'}{n}\right)\Big\{C_x^{\,2}\left(-2t_1^2g+t_1gK_{yx}\right)$$
$$+C_z^2\left(-t_1t_2g^2K_{xz}-t_2^2g^2+2t_1t_2g^2K_{xz}-t_2gK_{yz}\right)\Big\}$$

$$Bias(\bar{y}_{3d})=\bar{Y}\left(\frac{1-f}{n}\right)C_z^{\,2}\left(-\lambda_1\lambda_2g^2K_{xz}+\lambda_2^{\,2}g+\lambda_2gK_{yz}\right)+\bar{Y}\left(\frac{1-f'}{n}\right)\Big\{C_x^{\,2}\left(-2\lambda_1^2g+\lambda_1gK_{yx}\right)$$

$$+C_z^2\left(-\lambda_1\lambda_2 g^2 K_{xz} - \lambda_2^2 g^2 + 2\lambda_1\lambda_2 g^2 K_{xz} - \lambda_2 g K_{yz}\right)\Big\}$$

$$Bias\left(\bar{y}_{4d}\right) = \bar{Y}\left(\frac{f'-f}{n}\right)\left[\theta_1 C_x^2\left(\frac{\alpha\theta_1}{2} + \frac{\alpha^2\theta_1}{2} - \alpha K_{yx}\right) - \theta_2\alpha\left(\alpha\theta_1 C_{xz} + \alpha C_{yz}\right)\right]$$

$$-\bar{Y}\left(\frac{2-f-f'}{2n}\right)\left[\alpha\theta_2^2 C_z^2\left(1-\alpha\right)\right]$$

$$Bias\left(\bar{y}_{5d}\right) = \bar{Y}\left(\frac{f'-f}{n}\right)\left[t_1 C_x^2\left(\frac{\alpha t_1}{2} + \frac{\alpha^2 t_1}{2} - \alpha K_{yx}\right) - t_2\alpha\left(\alpha t_1 C_{xz} + \alpha C_{yz}\right)\right]$$

$$-\bar{Y}\left(\frac{2-f-f'}{2n}\right)\left[\alpha t_2^2 C_z^2\left(1-\alpha\right)\right]$$

และ

$$Bias\left(\bar{y}_{6d}\right) = \bar{Y}\left(\frac{f'-f}{n}\right)\left[\lambda_1 C_x^2\left(\frac{\alpha\lambda_1}{2} + \frac{\alpha^2\lambda_1}{2} - \alpha K_{yx}\right) - \lambda_2\alpha\left(\alpha\lambda_1 C_{xz} + \alpha C_{yz}\right)\right]$$

$$-\bar{Y}\left(\frac{2-f-f'}{2n}\right)\left[\alpha\lambda_2^2 C_z^2\left(1-\alpha\right)\right]$$

เมื่อ $\theta_1 = \dfrac{\bar{X}}{\bar{X}+C_x}$, $\theta_2 = \dfrac{\bar{Z}}{\bar{Z}+C_z}$, $C_{yx} = \rho_{yx} C_y C_x$, $C_{yz} = \rho_{yz} C_y C_z$, $g = \dfrac{n_1}{N-n_1}$

$$K_{xz} = \rho_{xz}\frac{C_x}{C_z},\ K_{yz} = \rho_{yz}\frac{C_y}{C_z},\ K_{yx} = \rho_{yx}\frac{C_y}{C_x},\ C_y^2 = \frac{S_y^2}{\bar{Y}^2},\ C_x^2 = \frac{S_x^2}{\bar{X}^2},\ C_z^2 = \frac{S_z^2}{\bar{Z}^2},$$

$t_1 = \dfrac{\bar{X}}{\bar{X}+\beta_2(x)}$, $t_2 = \dfrac{\bar{Z}}{\bar{Z}+\beta_2(z)}$, $\lambda_1 = \dfrac{\bar{X}}{\bar{X}+\rho_{xz}}$ และ $\lambda_2 = \dfrac{\bar{Z}}{\bar{Z}+\rho_{xz}}$

ความเอนเอียงของตัวประมาณที่นำเสนอในกรณีที่ 2

$$Bias\left(\bar{y}_{1d}\right) = \bar{Y}\left(\frac{1-f}{n}\right)C_z^{\,2}\left(-\theta_1\theta_2 g^2 K_{xz} + \theta_2^{\,2} g + \theta_2 g K_{yz}\right)$$

$$+\bar{Y}\left(\frac{1-f'}{n}\right)\left[-\theta_1^2 g C_x^{\,2} + C_z^2\left(-\theta_1\theta_2 g^2 K_{xz} + \theta_2^2 g^2 + \theta_2^2 g\right)\right]$$

$$Bias\left(\bar{y}_{2d}\right) = \bar{Y}\left(\frac{1-f}{n}\right)C_z^{\,2}\left(-t_1 t_2 g^2 K_{xz} + t_2^{\,2} g + t_2 g K_{yz}\right)$$

$$+\bar{Y}\left(\frac{1-f'}{n}\right)\left[-t_1^2 g C_x^{\,2} + C_z^2\left(-t_1 t_2 g^2 K_{xz} + t_2^2 g^2 + t_2^2 g\right)\right]$$

$$Bias\left(\bar{y}_{3d}\right) = \bar{Y}\left(\frac{1-f}{n}\right)C_z^{\,2}\left(-\lambda_1\lambda_2 g^2 K_{xz} + \lambda_2^{\,2} g + \lambda_2 g K_{yz}\right)$$

$$+\bar{Y}\left(\frac{1-f'}{n}\right)\left[-\lambda_1^{\,2} g C_x^{\,2} + C_z^2\left(-\lambda_1\lambda_2 g^2 K_{xz} + \lambda_2^2 g^2 + \lambda_2^2 g\right)\right]$$

$$Bias\left(\bar{y}_{4d}\right)=\bar{Y}\left(\frac{f'-f}{n}\right)\left[\frac{\alpha}{2}\left(\theta_1^2C_x^2-\theta_2^2C_z^2\right)\right]+\bar{Y}\left(\frac{1-f}{n}\right)\left[\alpha\left(\theta_2C_{yz}-\theta_1C_{yx}\right)\right]$$
$$+\bar{Y}\left(\frac{2-f-f'}{n}\right)\left\{\frac{\alpha^2}{2}\left[\theta_1^2C_x^2+\theta_2C_z^2\left(\theta_2-2\theta_1K_{xz}\right)\right]\right\}$$
$$Bias\left(\bar{y}_{5d}\right)=\bar{Y}\left(\frac{f'-f}{n}\right)\left[\frac{\alpha}{2}\left(t_1^2C_x^2-t_2^2C_z^2\right)\right]+\bar{Y}\left(\frac{1-f}{n}\right)\left[\alpha\left(t_2C_{yz}-t_1C_{yx}\right)\right]$$
$$+\bar{Y}\left(\frac{2-f-f'}{n}\right)\left\{\frac{\alpha^2}{2}\left[t_1^2C_x^2+t_2C_z^2\left(t_2-2t_1K_{xz}\right)\right]\right\}$$

และ

$$Bias\left(\bar{y}_{6d}\right)=\bar{Y}\left(\frac{f'-f}{n}\right)\left[\frac{\alpha}{2}\left(\lambda_1^2C_x^2-\lambda_2^2C_z^2\right)\right]+\bar{Y}\left(\frac{1-f}{n}\right)\left[\alpha\left(\lambda_2C_{yz}-\lambda_1C_{yx}\right)\right]$$
$$+\bar{Y}\left(\frac{2-f-f'}{n}\right)\left\{\frac{\alpha^2}{2}\left[\lambda_1^2C_x^2+\lambda_2C_z^2\left(\lambda_2-2\lambda_1K_{xz}\right)\right]\right\}$$

ความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณที่นำเสนอในกรณีที่ 1 คือ

$$MSE\left(\bar{y}_{1d}\right)=\bar{Y}^2\left(\frac{1-f}{n}\right)C_y^2$$
$$+\bar{Y}^2\left(\frac{f'-f}{n}\right)\left[C_z^{\,2}\left(2\theta_2gK_{yz}+\theta_2^{\,2}g^2-2\theta_1\theta_2g^2K_{xz}\right)+C_x^{\,2}\left(\theta_1^{\,2}g^2-2\theta_1gK_{yx}\right)\right]$$
$$MSE\left(\bar{y}_{2d}\right)=\bar{Y}^2\left(\frac{1-f}{n}\right)C_y^2$$
$$+\bar{Y}^2\left(\frac{f'-f}{n}\right)\left[C_z^{\,2}\left(2t_2gK_{yz}+t_2^{\,2}g^2-2t_1t_2g^2K_{xz}\right)+C_x^{\,2}\left(t_1^{\,2}g^2-2t_1gK_{yx}\right)\right]$$
$$MSE\left(\bar{y}_{3d}\right)=\bar{Y}^2\left(\frac{1-f}{n}\right)C_y^2$$
$$+\bar{Y}^2\left(\frac{f'-f}{n}\right)\left[C_z^{\,2}\left(2\lambda_2gK_{yz}+\lambda_2^{\,2}g^2-2\lambda_1\lambda_2g^2K_{xz}\right)+C_x^{\,2}\left(\lambda_1^{\,2}g^2-2\lambda_1gK_{yx}\right)\right]$$
$$MSE\left(\bar{y}_{4d}\right)=\bar{Y}^2\left(\frac{1-f}{n}\right)C_y^2$$
$$+\bar{Y}^2\left(\frac{f'-f}{n}\right)\left[\theta_1C_x^2\left(\alpha^2\theta_1-2\alpha K_{yx}\right)+\theta_2C_z^2\left(\alpha^2\theta_2+2\alpha K_{yz}-2\alpha^2\theta_1K_{xz}\right)\right]$$
$$MSE\left(\bar{y}_{5d}\right)=\bar{Y}^2\left(\frac{1-f}{n}\right)C_y^2$$
$$+\bar{Y}^2\left(\frac{f'-f}{n}\right)\left[t_1C_x^2\left(\alpha^2t_1-2\alpha K_{yx}\right)+t_2C_z^2\left(\alpha^2t_2+2\alpha K_{yz}-2\alpha^2t_1K_{xz}\right)\right]$$

และ

$$MSE\left(\bar{y}_{6d}\right)=\bar{Y}^2\left(\frac{1-f}{n}\right)C_y^2$$
$$+\bar{Y}^2\left(\frac{f'-f}{n}\right)\left[\lambda_1 C_x^2\left(\alpha^2\lambda_1-2\alpha K_{yx}\right)+\lambda_2 C_z^2\left(\alpha^2\lambda_2+2\alpha K_{yz}-2\alpha^2\lambda_1 K_{xz}\right)\right]$$

ความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณที่นำเสนอในกรณีที่ 2 คือ

$$MSE\left(\bar{y}_{1d}\right)=\bar{Y}^2\left(\frac{1-f}{n}\right)\left[C_y^2+C_z^{\,2}\left(2\theta_2 gK_{yz}+\theta_2^{\,2}g^2-2\theta_1\theta_2 g^2K_{xz}\right)+C_x^{\,2}\left(\theta_1^{\,2}g^2-2\theta_1 gK_{yx}\right)\right]$$
$$+\bar{Y}^2\left(\frac{1-f'}{n}\right)\left[\theta_1^2 g^2 C_x^{\,2}+C_z^{\,2}\left(\theta_2^{\,2}g^2-2\theta_1\theta_2 g^2K_{xz}\right)\right]$$
$$MSE\left(\bar{y}_{2d}\right)=\bar{Y}^2\left(\frac{1-f}{n}\right)\left[C_y^2+C_z^{\,2}\left(2t_2 gK_{yz}+t_2^{\,2}g^2-2t_1t_2 g^2K_{xz}\right)+C_x^{\,2}\left(t_1^{\,2}g^2-2t_1 gK_{yx}\right)\right]$$
$$+\bar{Y}^2\left(\frac{1-f'}{n}\right)\left[t_1^{\,2}g^2C_x^{\,2}+C_z^{\,2}\left(t_2^{\,2}g^2-2t_1t_2 g^2K_{xz}\right)\right]$$
$$MSE\left(\bar{y}_{3d}\right)=\bar{Y}^2\left(\frac{1-f}{n}\right)\left[C_y^2+C_z^{\,2}\left(2\lambda_2 gK_{yz}+\lambda_2^{\,2}g^2-2\lambda_1\lambda_2 g^2K_{xz}\right)+C_x^{\,2}\left(\lambda_1^{\,2}g^2-2\lambda_1 gK_{yx}\right)\right]$$
$$+\bar{Y}^2\left(\frac{1-f'}{n}\right)\left[\lambda_1^{\,2}g^2C_x^{\,2}+C_z^{\,2}\left(\lambda_2^{\,2}g^2-2\lambda_1\lambda_2 g^2K_{xz}\right)\right]$$
$$MSE\left(\bar{y}_{4d}\right)=\bar{Y}^2\left(\frac{1-f}{n}\right)\left(C_y^2-2\alpha\theta_1 C_{yx}+2\alpha\theta_2 C_{yz}\right)$$
$$+\bar{Y}^2\left(\frac{2-f'-f}{n}\right)\left[\alpha^2\theta_1^2C_x^2+\alpha^2\theta_2 C_z^2\left(\theta_2-2\theta_1 K_{xz}\right)\right]$$
$$MSE\left(\bar{y}_{5d}\right)=\bar{Y}^2\left(\frac{1-f}{n}\right)\left(C_y^2-2\alpha t_1 C_{yx}+2\alpha t_2 C_{yz}\right)$$
$$+\bar{Y}^2\left(\frac{2-f'-f}{n}\right)\left[\alpha^2 t_1^2C_x^2+\alpha^2 t_2 C_z^2\left(t_2-2t_1 K_{xz}\right)\right]$$

และ

$$MSE\left(\bar{y}_{6d}\right)=\bar{Y}^2\left(\frac{1-f}{n}\right)\left(C_y^2-2\alpha\lambda_1 C_{yx}+2\alpha\lambda_2 C_{yz}\right)$$
$$+\bar{Y}^2\left(\frac{2-f'-f}{n}\right)\left[\alpha^2\lambda_1^2C_x^2+\alpha^2\lambda_2 C_z^2\left(\lambda_2-2\lambda_1 K_{xz}\right)\right]$$

สำหรับตัวประมาณที่นำเสนอ $\bar{y}_{4d}$, $\bar{y}_{5d}$ และ $\bar{y}_{6d}$ ในทั้ง 2 กรณี ทำการหาค่า α ที่ทำให้ความคลาดเคลื่อนกำลังสองเฉลี่ยของ $\bar{y}_{4d}$, $\bar{y}_{5d}$ และ $\bar{y}_{6d}$ มีค่าต่ำสุด โดยพบว่า

กรณีที่ 1

ค่า α ที่ทำให้ความคลาดเคลื่อนกำลังสองเฉลี่ยของ $\bar{y}_{4d}$ มีค่าต่ำสุด คือ $\alpha=\dfrac{\theta_1 C_x^2K_{yx}+\theta_2 C_z^2K_{yz}}{\theta_1^2C_x^2+\theta_2 C_z^2\left(\theta_2-2\theta_1 K_{xz}\right)}$

ค่า α ที่ทำให้ความคลาดเคลื่อนกำลังสองเฉลี่ยของ $\bar{y}_{5d}$ มีค่าต่ำสุด คือ $\alpha = \dfrac{t_1 C_x^2 K_{yx} + t_2 C_z^2 K_{yz}}{t_1^2 C_x^2 + t_2 C_z^2 \left(t_2 - 2t_1 K_{xz}\right)}$

ค่า α ที่ทำให้ความคลาดเคลื่อนกำลังสองเฉลี่ยของ $\bar{y}_{6d}$ มีค่าต่ำสุด คือ $\alpha = \dfrac{\lambda_1 C_x^2 K_{yx} + \lambda_2 C_z^2 K_{yz}}{\lambda_1^2 C_x^2 + \lambda_2 C_z^2 \left(\lambda_2 - 2\lambda_1 K_{xz}\right)}$

กรณีที่ 2

ค่า α ที่ทำให้ความคลาดเคลื่อนกำลังสองเฉลี่ยของ $\bar{y}_{4d}$ มีค่าต่ำสุด คือ $\alpha = \dfrac{\theta_1 M_{yx} - \theta_2 M_{yz}}{f_1\left[\theta_1^2 C_x^2 + \theta_2 C_z^2 \left(\theta_2 - 2\theta_1 K_{xz}\right)\right]}$

ค่า α ที่ทำให้ความคลาดเคลื่อนกำลังสองเฉลี่ยของ $\bar{y}_{5d}$ มีค่าต่ำสุด คือ $\alpha = \dfrac{t_1 M_{yx} - t_2 M_{yz}}{f_1\left[t_1^2 C_x^2 + t_2 C_z^2 \left(t_2 - 2t_1 K_{xz}\right)\right]}$

ค่า α ที่ทำให้ความคลาดเคลื่อนกำลังสองเฉลี่ยของ $\bar{y}_{6d}$ มีค่าต่ำสุด คือ

$$\alpha = \frac{\lambda_1 M_{yx} - \lambda_2 M_{yz}}{f_1\left[\lambda_1^2 C_x^2 + \lambda_2 C_z^2 \left(\lambda_2 - 2\lambda_1 K_{xz}\right)\right]}$$

โดย $M_{yx} = \rho_{yx} C_y C_x$ และ $M_{yz} = \rho_{yz} C_y C_z$

ตัวอย่างเชิงตัวเลข

ทำการแสดงตัวอย่างเชิงตัวเลขเพื่อทำการสนับสนุนผลของการศึกษาเชิงทฤษฎี โดยใช้ข้อมูลจากตัวแปรช่วย x ที่มีความสัมพันธ์ทิศทางเดียวกันกับตัวแปรที่ต้องการศึกษา y และข้อมูลจากตัวแปรช่วย z ที่มีความสัมพันธ์ทิศทางตรงข้ามกับตัวแปรที่ต้องการศึกษา y ซึ่งข้อมูลที่นำมาแสดงมี 2 ชุดดังมีรายละเอียดและผลลัพธ์ของการศึกษา ดังต่อไปนี้

1. ข้อมูลจาก Steel และ Torrie ในปี 1960 [9] มีข้อมูลของอัตราการเผาไหม้ของใบไม้เป็นตัวแปรที่ต้องการศึกษา y มีตัวแปรอิสระ x และ z เป็นร้อยละของโพสแทสเซียม และร้อยละของคลอรีน ตามลำดับ โดยข้อมูลมีสารสนเทศดังต่อไปนี้

N = 30, n_1 = 14, n = 6, $\bar{Y}$ = 0.6860, C_y = 0.4803, $\beta_2(x)$ = 1.56, ρ_{xy} = 0.1794, $\bar{X}$ = 4.6537, C_x = 0.2295, $\beta_2(z)$ = 1.40, ρ_{yz} = -0.4996, $\bar{Z}$ = 0.8077, C_z = 0.7493, ρ_{xz} = 0.4074

2. ข้อมูลจาก Thomas ในปี 1990 [10] มีข้อมูลอัตราการตายต่อหนึ่งพันครัวเรือน เป็นตัวแปรที่ต้องการศึกษา y มีตัวแปรอิสระ x และ z เป็นจำนวนโรงพยาบาลที่พร้อมให้บริการต่อหนึ่งแสนครัวเรือน และความหนาแน่นของประชากรต่อหนึ่งตารางไมล์ ตามลำดับ โดยมีข้อมูลสารสนเทศ ต่อไปนี้

N = 53, n_1 = 26, n = 11, $\bar{Y}$ = 9.3057, C_y = 0.1787, $\beta_2(x)$ = 3.8800, ρ_{xy} = 0.1110, $\bar{X}$ = 589.7925, C_x = 0.5640, $\beta_2(z)$ = 2.9860, ρ_{yz} = -0.2780, $\bar{Z}$ = 110.6415, C_z = 0.4265, ρ_{xz} = 0.1870

โดยทำการเปรียบเทียบประสิทธิภาพของตัวประมาณที่นำเสนอ $\bar{y}_{1d}$, $\bar{y}_{2d}$, $\bar{y}_{3d}$ กับตัวประมาณอัตราส่วนร่วมกับผลคูณของ Singh ในปี 1967 [1] และตัวประมาณอัตราส่วนร่วมกับผลคูณของ Singh และคณะในปี 2005 [3] และทำการเปรียบเทียบประสิทธิภาพของตัวประมาณที่นำเสนอ $\bar{y}_{4d}$, $\bar{y}_{5d}$, $\bar{y}_{6d}$ กับตัวประมาณอัตราส่วนร่วมกับผลคูณของ Chanu และ Singh ในปี 2014 [8] โดยพิจารณาจากความคลาดเคลื่อนกำลังสองเฉลี่ย ซึ่งตัวประมาณที่นำเสนอจะมีประสิทธิภาพมากกว่าตัวประมาณอื่นๆก็ต่อเมื่อ $MSE(\hat{\theta}_0) < MSE(\hat{\theta})$ โดยที่ $\hat{\theta}_0$ เป็นตัวประมาณที่นำเสนอ และ $\hat{\theta}$ เป็นตัวประมาณแบบดั้งเดิมของ θ แสดงผลได้ดังตารางที่ 1 และ 3 จากนั้นทำการหาร้อยละประสิทธิภาพ (Relative Efficiency: RE) ของตัวประมาณที่นำเสนอเทียบกับตัวประมาณแบบดั้งเดิม ซึ่งมีรูปแบบ คือ

$$RE(\hat{\theta}, \hat{\theta}_0) = \left(\frac{MSE(\hat{\theta})}{MSE(\hat{\theta}_0)}\right) \times 100\%$$

ซึ่งแสดงผลได้ดังตารางที่ 2 และ 4

ตารางที่ 1 ความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณ $\bar{y}_{RP}$ $\bar{y}_{RP}^{*}$ $\bar{y}_{1d}$ $\bar{y}_{2d}$ และ $\bar{y}_{3d}$

ข้อมูลชุดที่	ตัวประมาณ							
	$\bar{y}_{RP}$	$\bar{y}_{RP}^{*}$	$\bar{y}_{1d}$		$\bar{y}_{2d}$		$\bar{y}_{3d}$	
			กรณีที่ 1	กรณีที่ 2	กรณีที่ 1	กรณีที่ 2	กรณีที่ 1	กรณีที่ 2
1	0.07521	0.02143	0.01221	0.00962	0.01270	0.00992	0.01236	0.00987
2	3.90524	1.79258	0.21967	1.95872	0.21492	1.92558	0.22012	1.96225

ตารางที่ 2 ร้อยละประสิทธิภาพของตัวประมาณ $\bar{y}_{RP}^{*}$ $\bar{y}_{1d}$ $\bar{y}_{2d}$ และ $\bar{y}_{3d}$ เทียบกับ $\bar{y}_{RP}$

ข้อมูลชุดที่	ตัวประมาณ							
	$\bar{y}_{RP}$	$\bar{y}_{RP}^{*}$	$\bar{y}_{1d}$		$\bar{y}_{2d}$		$\bar{y}_{3d}$	
			กรณีที่ 1	กรณีที่ 2	กรณีที่ 1	กรณีที่ 2	กรณีที่ 1	กรณีที่ 2
1	100	350.92435	615.99254	781.44948	592.32767	757.76812	608.22966	762.18547
2	100	217.85594	1777.73898	199.37749	1817.04483	202.80886	1774.11535	199.01854

ตารางที่ 3 ความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณ $\bar{y}_{RP}$ $\bar{y}_{RP}^{w(d)}$ $\bar{y}_{4d}$ $\bar{y}_{5d}$ และ $\bar{y}_{6d}$

ข้อมูลชุดที่	ตัวประมาณ								
	$\bar{y}_{RP}$	$\bar{y}_{RP}^{w(d)}$		$\bar{y}_{4d}$		$\bar{y}_{5d}$		$\bar{y}_{6d}$	
		กรณีที่ 1	กรณีที่ 2	กรณีที่ 1	กรณีที่ 2	กรณีที่ 1	กรณีที่ 2	กรณีที่ 1	กรณีที่ 2
1	0.07521	0.01071	0.01530	0.01806	0.01435	0.02008	0.01431	0.01921	0.01428
2	3.90524	0.18762	0.20134	0.20327	0.19627	0.20320	0.19632	0.20328	0.19627

ตารางที่ 4 ร้อยละประสิทธิภาพของตัวประมาณ $\bar{y}_{RP}^{w(d)}$ $\bar{y}_{4d}$ $\bar{y}_{5d}$ และ $\bar{y}_{6d}$ เทียบกับ $\bar{y}_{RP}$

ข้อมูลชุดที่	ตัวประมาณ								
	$\bar{y}_{RP}$	$\bar{y}_{RP}^{w(d)}$		$\bar{y}_{4d}$		$\bar{y}_{5d}$		$\bar{y}_{6d}$	
		กรณีที่ 1	กรณีที่ 2	กรณีที่ 1	กรณีที่ 2	กรณีที่ 1	กรณีที่ 2	กรณีที่ 1	กรณีที่ 2
1	100	702.30007	491.46987	416.41643	523.97846	374.46877	525.57418	391.54720	526.76717
2	100	2081.42678	1939.66268	1921.19780	1989.71340	1921.86186	1989.22816	1921.13212	1989.76069

จากตารางที่ 1 และ 2 พบว่า ตัวประมาณ $\bar{y}_{1d}$ มีประสิทธิภาพมากกว่าตัวประมาณอื่นๆใน และมีความคลาดเคลื่อนกำลังสองเฉลี่ยน้อยที่สุดในข้อมูลชุดที่ 1 ทั้งกรณีที่ 1 และกรณีที่ 2 แต่ในข้อมูลชุดที่ 2 ตัวประมาณ $\bar{y}_{2d}$ มีประสิทธิภาพมากที่สุดและมีความคลาดเคลื่อนกำลังสองเฉลี่ยน้อยที่สุดในกรณีที่ 1 และตัวประมาณ $\bar{y}_{RP}^{*}$ มีประสิทธิภาพมากที่สุดและมีความคลาดเคลื่อนกำลังสองน้อยที่สุดในกรณีที่ 2 ตามลำดับ

จากตารางที่ 3 และ 4 พบว่า ในกรณีที่ 1 ตัวประมาณ $\bar{y}_{RP}^{w(d)}$ มีประสิทธิภาพมากกว่าตัวประมาณอื่นๆ และมีความคลาดเคลื่อนกำลังสองเฉลี่ยน้อยที่สุดในทุกชุดข้อมูล ส่วนในกรณีที่ 2 พบว่า ตัวประมาณ $\bar{y}_{6d}$ มีประสิทธิภาพมากที่สุดและมีความคลาดเคลื่อนกำลังสองเฉลี่ยน้อยที่สุดในทุกชุดข้อมูล

สำหรับการเปรียบเทียบความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณ $\bar{y}_{1d}$ $\bar{y}_{2d}$ $\bar{y}_{3d}$ $\bar{y}_{4d}$

$\bar{y}_{5d}$ และ $\bar{y}_{6d}$ ระหว่างกรณีที่ 1 และกรณีที่ 2 สามารถ แสดงได้ดังตารางที่ 5

ตารางที่ 5 เปรียบเทียบความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณ $\bar{y}_{1d}$ $\bar{y}_{2d}$ $\bar{y}_{3d}$ $\bar{y}_{4d}$ $\bar{y}_{5d}$ และ $\bar{y}_{6d}$ ระหว่างกรณีที่ 1 และกรณีที่ 2

ข้อมูลชุดที่	**1**		**2**	
ตัวประมาณ	**กรณีที่ 1**	**กรณีที่ 2**	**กรณีที่ 1**	**กรณีที่ 2**
$\bar{y}_{1d}$	0.01221	0.00962	0.21968	1.95872
$\bar{y}_{2d}$	0.01270	0.00992	0.21492	1.92558
$\bar{y}_{3d}$	0.01237	0.00987	0.22012	1.96225
$\bar{y}_{4d}$	0.01806	0.01435	0.20327	0.19627
$\bar{y}_{5d}$	0.02008	0.01431	0.20320	0.19632
$\bar{y}_{6d}$	0.01921	0.01428	0.20328	0.19627

จากตารางที่ 5 พบว่า ตัวประมาณนำเสนอ $\bar{y}_{1d}$ $\bar{y}_{2d}$ และ $\bar{y}_{3d}$ ในข้อมูลชุดที่ 1 ในกรณีที่ 1 มีความคลาดเคลื่อนกำลังสองเฉลี่ยน้อยกว่ากรณีที่ 2 แต่ในข้อมูลชุดที่ 2 ตัวประมาณที่นำเสนอในกรณีที่ 2 มีความคลาดเคลื่อนกำลังสองเฉลี่ยน้อยกว่ากรณีที่ 1 ส่วนตัวประมาณ $\bar{y}_{4d}$ $\bar{y}_{5d}$ และ $\bar{y}_{6d}$ ในกรณีที่ 2 มีความคลาดเคลื่อนกำลังสองเฉลี่ยน้อยกว่ากรณีที่ 1 ในทุกชุด ซึ่งแสดงว่าตัวประมาณที่นำเสนอ $\bar{y}_{1d}$ $\bar{y}_{2d}$ และ $\bar{y}_{3d}$ กรณีที่ 1 และกรณีที่ 2 มีประสิทธิภาพแตกต่างกันในบางชุดข้อมูล แต่ตัวประมาณ $\bar{y}_{4d}$ $\bar{y}_{5d}$ และ $\bar{y}_{6d}$ ในกรณีที่ 2 มีประสิทธิภาพมากกว่าตัวประมาณที่นำเสนอในกรณีที่ 1

สรุปและข้อเสนอแนะ

จากการศึกษาตัวประมาณอัตราส่วนร่วมกับผลคูณสำหรับการประมาณค่าเฉลี่ยของประชากร ภายใต้การเลือกตัวอย่างสุ่มแบบง่ายสองเฟสแบบไม่ใส่คืน โดยผู้วิจัยได้ทำการปรับปรุงตัวประมาณอัตราส่วนร่วมกับผลคูณของ Singh และคณะ ในปี 2005 [3] และตัวประมาณอัตราส่วนร่วมกับผลคูณของ Chanu และ Singh ในปี 2014 [8] ให้มีประสิทธิภาพมากยิ่งขึ้นโดยการนำ x_i และ z_i $;i=1,2,...,N$ ที่อยู่ในรูปของ $x_i^*=(1+g)\bar{x}_1-gx_i$ และ $z_i^*=(1+g)\bar{z}_1-gz_i$ มาใช้ร่วมกับสัมประสิทธิ์การแปรผันของตัวแปรช่วย x (C_x) สัมประสิทธิ์การแปรผันของตัวแปรช่วย z (C_z) สัมประสิทธิ์ความโด่งของตัวแปรช่วย x ($\beta_2(x)$) สัมประสิทธิ์ความโด่งของตัวแปรช่วย z ($\beta_2(z)$) และสัมประสิทธิ์สหสัมพันธ์ของตัวแปรช่วย x และ z (ρ_{xz}) ในการปรับปรุงตัวประมาณ ซึ่งตัวประมาณที่นำเสนอแบ่งออกเป็น 2 กรณี คือ กรณีที่ 1 หน่วยตัวอย่างสัมพันธ์กัน และกรณีที่ 2 หน่วยตัวอย่างอิสระกัน ซึ่งจากผลการศึกษาจะเห็นได้ว่าตัวประมาณที่นำเสนอ มีประสิทธิภาพมากกว่าตัวประมาณ $\bar{y}_{RP}^*$ และ $\bar{y}_{RP}^{w(d)}$ ในบางชุดข้อมูล และเมื่อทำการเปรียบเทียบประสิทธิภาพของตัวประมาณทั้ง 2 กรณี พบว่าตัวประมาณที่นำเสนอมีประสิทธิภาพมากกว่าตัวประมาณแบบดั้งเดิม ทั้งในกรณีตัวอย่างเป็นอิสระกัน และกรณีตัวอย่างสัมพันธ์กัน

เอกสารอ้างอิง

[1] Singh, M.P. 1967. “Ratio cum product method of estimation”, **Metrika**, 12(1), 34-42.

[2] Tracy, D.S.; Singh, H.P.; Singh, R. 1996. “An alternative to the ratio-cum-product estimator in sample surveys”, **Journal of Statistical Planning and Inference**, 53, 375-387.

[3] Singh, H.P.; Singh, R.; Espejo, M.R.; Pineda, D.; Nadarajah, S. 2005. "On the efficiency of a dual to ratio-cum-product estimators in sample surveys", **Mathematical Proceedings of Royal Irish Academy**, 105A(2), 51-56.

[4] Singh, H.P.; Tailor, R. 2005. ''Estimation of finite population mean using known correlation coefficient between auxiliary characters''. **Statistica**, (65)4, 407-418.

[5] Tailor, R.; Verma, M.R.; Sharma, B. 2010. "An alternative ratio-cum-product estimator of population mean using a coefficient of kurtosis for two auxiliary variables", **Data Science Journal**, 9, 93-99.

[6] Tailor, R.; Parmar, R.; Kim, J.M.; Tailor, R. 2011. ''Ratio-cum-product estimators of population mean using known population parameters of auxiliary variates'', **Communications of the Korean Statistical Society**. 18(2), 155–164.

[7] Tailor, R.; Tailor, R.; Parmar, R.; Kumar, M. 2012. ''Dual to ratio-cum-product estimator using known parameters of auxiliary variables'' **Journal of Reliability and Statistical Studies**, 5(1), 65-71.

[8] Chanu, W.W.; Singh, B.K. 2014. ''Improve class of ratio-cum-product estimators of finite population mean in two phase sampling'' **Global Journal of Science Frontier Research: F Mathematics and Decision Sciences**. 14(2), 68-81.

[9] Steel, R.; Torrie, H.J. *Principles and procedures of Statistics*; McGraw-Hill: New York, 1960.

[10] Thomas, G.S. *The Rating Guide to Life in America's Small Cities*. Prometheus Books: New York, 1990