ตัวประมาณค่าเฉลี่ยประชากรในการเลือกตัวอย่างแบบแบ่งชั้นภูมิอย่างง่าย : วิธีของซิลส์

Estimator of Population Mean in Stratified Random Sampling : Searls Approach

นฤพล วงศ์เจริญสันติ*, จิราวัลย์ จิตรถเวช และ วิชิต หล่อจีระชุณห์กุล

Narupon Wongcharoensanti*, Jirawan Jitthavech and Vichit Lorchirachoonkul

คณะสถิติประยุกต์ สถาบันบัณฑิตพัฒนบริหารศาสตร์

School of Applied Statistics, National Institute of Development Administration

Received : 18 May 2018

Accepted : 5 September 2018

Published online : 17 September 2018

บทคัดย่อ

การศึกษานี้ได้นำเสนอตัวประมาณค่าเฉลี่ยประชากรใหม่ 2 ตัวที่ใช้ค่าสัมประสิทธิ์การแปรผัน ในการเลือกตัวอย่างแบบแบ่งชั้นภูมิอย่างง่ายไม่คืนที่ ตัวประมาณตัวที่หนึ่งเป็นตัวประมาณที่ไม่ใช้ตัวแปรช่วย พัฒนาจากตัวประมาณของซิลส์ และตัวประมาณตัวที่สองเป็นตัวประมาณอัตราส่วนที่พัฒนาจากตัวประมาณอัตราส่วนของ Jitthavech and Lorchirachoonkul พร้อมด้วยเงื่อนไขที่ตัวประมาณตัวที่หนึ่งมีประสิทธิภาพสูงกว่าตัวประมาณดั้งเดิมที่ไม่ใช้ตัวแปรช่วย และเงื่อนไขที่ตัวประมาณตัวที่สองมีประสิทธิภาพสูงกว่าตัวประมาณอัตราส่วนดั้งเดิมแบบแยก ผลการจำลองยืนยันความสอดคล้องของเงื่อนไขทั้งสองที่ได้วิเคราะห์ไว้ นอกจากนี้ ยังได้เปรียบเทียบกับตัวประมาณอัตราส่วนที่พัฒนาจากตัวประมาณอัตราส่วนของ Sisodia and Dwivedi ซึ่งสรุปจากผลการจำลองได้ว่าตัวประมาณตัวที่สองมีประสิทธิภาพสูงกว่า และมีความเอนเอียงต่ำกว่าด้วย

คำสำคัญ: ตัวประมาณค่าของ Searls, ตัวประมาณอัตราส่วนของ Jitthavech and Lorchirachoonkul, การเลือกตัวอย่างแบบแบ่งชั้นภูมิอย่างง่ายไม่คืนที่, ประสิทธิภาพสัมพัทธ์

Abstract

The study proposes two new population mean estimators using the coefficient of variation in stratified random sampling without replacement. The first estimator without an auxiliary variable is developed from Searls estimator and the second estimator is a ratio estimator developed from Jitthavech and Lorchirachoonkul ratio estimator. The conditions for the efficiency of the first proposed estimator higher than the conventional estimator without auxiliary variable and for the efficiency of the second proposed estimator higher than the conventional separate ratio estimator are also analyzed. The simulation confirms the consistency of the two conditions. Moreover, it can be concluded from the simulation that the efficiency of the second estimator is higher than the separate ratio estimator developed from Sisodia and Dwivedi ratio estimator with the lower biases.

Keywords: Searls estimator, Jitthavech and Lorchirachoonkul ratio estimator, stratified random sampling without replacement, relative efficiency.

***Corresponding author. E-mail:** narupon.won@gmail.com

บทนำ

การเลือกตัวอย่างแบบแบ่งเป็นชั้นภูมิเป็นวิธีการที่ใช้กันอย่างแพร่หลายในงานวิจัยที่มีการสำรวจขนาดใหญ่ และต้องการสารสนเทศในรายชั้นภูมิ การเลือกตัวอย่างแบบแบ่งชั้นภูมิ (Stratified sampling) เป็นการเลือกตัวอย่างโดยแบ่งประชากรออกเป็นส่วนๆ แต่ละส่วนเรียกว่าชั้นภูมิ (Stratum) หน่วยตัวอย่างจะอยู่ในชั้นภูมิใดชั้นภูมิหนึ่งไม่ซ้ำซ้อนกัน หลักในการแบ่งชั้นภูมิที่ดีต้องให้หน่วยตัวอย่างในแต่ละชั้นภูมิมีลักษณะของตัวแปรที่สนใจคล้ายคลึงกัน และมีลักษณะของตัว แปรที่สนใจแตกต่างกันระหว่างชั้นภูมิ ในทางปฏิบัติผู้วิจัยมักไม่มีข้อมูลเกี่ยวกับตัวแปรสนใจที่กำลังจะศึกษาเพื่อใช้ในการจัดแยกหน่วยตัวอย่างไปตามชั้นภูมิ ซึ่งอาจใช้ข้อมูลในอดีตของตัวแปรที่สนใจหรืออาจใช้ตัวแปรช่วยที่มีความสัมพันธ์กับตัวแปรที่สนใจมาช่วยในการแบ่งชั้นภูมิ เมื่อแบ่งประชากรออกเป็นชั้นภูมิตามจำนวนของชั้นภูมิที่กำหนดไว้แล้ว ทำการสุ่มหน่วยตัวอย่างจากทุกชั้นภูมิอย่างเป็นอิสระกัน แต่ละชั้นภูมิอาจใช้วิธีการเลือกตัวอย่างที่แตกต่างกันหรือเหมือนกันได้ การสุ่มหน่วยตัวอย่างโดยการแบ่งประชากรออกเป็นชั้นภูมิ ทำให้สามารถเก็บคุณลักษณะที่สำคัญของประชากรไว้ในตัวอย่างได้สมบูรณ์กว่าการสุ่มหน่วยตัวอย่างแบบง่าย โดยเฉพาะคุณลักษณะของประชากรที่มีสัดส่วนน้อย

การประมาณค่าเฉลี่ยประชากรในการเลือกตัวอย่างแบบแบ่งชั้นภูมิ สามารถกระทำได้หลายวิธี การศึกษานี้ครอบคลุมตัวประมาณที่ไม่ใช้ตัวแปรช่วยและตัวประมาณอัตราส่วนซึ่งมีอยู่สองแนวทาง คือการประมาณค่าแบบรวม (Combined estimate) และการประมาณค่าแบบแยก (Separate estimate) โดยทั่วๆ ไป ถ้าขนาดตัวอย่างในชั้นภูมิใหญ่ การประมาณค่าแบบแยก มีความแม่นยำสูงกว่าการประมาณค่าแบบรวม ยกเว้นในกรณีที่อัตราส่วนในแต่ละชั้นภูมิมีค่าเท่ากัน (Cochran, 1977) การศึกษาเพื่อเพิ่มประสิทธิภาพของตัวประมาณค่าแบบอัตราส่วน ได้รับความสนใจในหมู่นักวิชาการมากมาย โดยเสนอตัวประมาณค่าที่ใช้สัมประสิทธิ์ความแปรผันและ/หรือความโด่งของตัวแปรช่วย (Kadilar and Cingi, 2003; Koyuncu and Kadilar, 2009; Tailor and Lone, 2014; Sangngam, 2014) ตัวประมาณค่าในการเลือกตัวอย่างแบบแบ่งชั้นภูมิที่มีการศึกษา เป็นตัวประมาณที่พัฒนาจากตัวประมาณค่าในการเลือกตัวอย่างสุ่มแบบง่าย เช่น ตัวประมาณของ Sisodia and Dwivedi (1981); Upadhyaya and Singh (1999); Khoshnevisan *et al*. (2007) เป็นต้น แต่ยังไม่มีการศึกษาตัวประมาณค่าแบบอัตราส่วนในการเลือกตัวอย่างแบบแบ่งชั้นภูมิ ที่พัฒนาจากตัวประมาณของซิลส์ ซึ่งเป็นตัวประมาณที่มีประสิทธิภาพสูงสุดในการเลือกตัวอย่างสุ่มแบบง่าย (Searls, 1964; Jitthavech and Lorchirachoonkul, 2013)

ดังนั้น การศึกษานี้ จึงมุ่งพัฒนาตัวประมาณที่ไม่มีตัวแปรช่วยและตัวประมาณแบบอัตราส่วนในการเลือกตัวอย่างแบบแบ่งชั้นภูมิโดยพัฒนาต่อจากการศึกษาของ Jitthavech and Lorchirachoonkul (2013) พร้อมทั้งศึกษาคุณสมบัติและเงื่อนไขของตัวประมาณที่พัฒนาขึ้น ที่มีประสิทธิภาพสูงกว่าตัวประมาณแบบดั้งเดิม ตลอดจนใช้การจำลองข้อมูล เพื่อศึกษาความสอดคล้องกับเงื่อนไขที่พัฒนาขึ้น

การทบทวนวรรณกรรม

การทบทวนวรรณกรรม จะเน้นวรรณกรรมตัวประมาณค่าเฉลี่ยประชากรแบบไม่ใช้ตัวแปรช่วยและแบบอัตราส่วนในการเลือกตัวอย่างแบบแบ่งชั้นภูมิอย่างง่าย (Stratified random sampling) และวรรณกรรมที่เกี่ยวข้อง ตัวประมาณค่าเฉลี่ยประชากรแบบดั้งเดิมที่ไม่ใช้ตัวแปรช่วย (Cochran, 1977) สามารถเขียนได้ดังนี้

$$\bar{y}_{st} = \sum_{h=1}^{k} w_h \bar{y}_h \tag{1}$$

เมื่อ $\bar{y}_h$ เป็นค่าเฉลี่ยตัวอย่างของตัวแปรที่สนใจในชั้นภูมิ h, $h=1,2,\ldots,k$, k เป็นจำนวนชั้นภูมิ, $w_h = N_h/N$, N เป็นขนาดประชากร และ N_h เป็นจำนวนหน่วยในชั้นภูมิ h

ตัวประมาณ $\bar{y}_{st}$ เป็นตัวประมาณไม่เอนเอียงของค่าเฉลี่ยประชากร ที่มีค่าคลาดเคลื่อนกำลังสองเฉลี่ยเท่ากับ

$$\mathrm{MSE}\left(\bar{y}_{st}\right) = \sum_{h=1}^{k} w_h^2 \gamma_h C_{yh}^2 \bar{Y}_h^2 \tag{2}$$

เมื่อ C_{yh} เป็นสัมประสิทธิ์การแปรผันประชากรของตัวแปรที่สนใจในชั้นภูมิ h, $\bar{Y}_h$ เป็นค่าเฉลี่ยประชากรของตัวแปรที่สนใจในชั้นภูมิ h, $\gamma_h = 1/n_h - 1/N_h$ และ n_h เป็นจำนวนหน่วยตัวอย่างในชั้นภูมิ h ซึ่งในทางปฏิบัติ พารามิเตอร์ที่เกี่ยวกับตัวแปรที่สนใจ อาจต้องใช้ค่าประมาณในการคำนวณแทนค่าจริง

เมื่อทราบสารสนเทศของตัวแปรช่วยในทุกชั้นภูมิ อาจใช้การประมาณค่าแบบอัตราส่วน ซึ่งตัวประมาณค่าเฉลี่ยประชากรแบบอัตราส่วนเดิมที่นิยมใช้ในการเลือกตัวอย่างแบบแบ่งชั้นภูมิอย่างง่ายโดยไม่คืนที่มี 2 ตัว (Cochran, 1977) คือ ตัวประมาณแบบอัตราส่วนรวม (Combined ratio estimator) และตัวประมาณแบบอัตราส่วนแยก (Separate ratio estimator) ซึ่งเขียนได้ตามลำดับดังนี้

$$\bar{y}_{cr} = \frac{\bar{y}_{st}}{\bar{x}_{st}} \bar{X} \tag{3}$$

$$\bar{y}_{sr} = \sum_{h=1}^{k} w_h \frac{\bar{y}_h}{\bar{x}_h} \bar{X}_h \tag{4}$$

เมื่อ $\bar{x}_{st} = \sum_{h=1}^{k} w_h \bar{x}_h$, $\bar{x}_h$ เป็นค่าเฉลี่ยตัวอย่างของตัวแปรช่วยในชั้นภูมิ h, $\bar{X}_h$ และ $\bar{X}$ เป็นค่าเฉลี่ยประชากรของตัวแปรช่วยในชั้นภูมิ h และทั้งประชากรตามลำดับ

ความเอนเอียง และค่าคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณทั้งสอง เขียนได้ดังนี้

$$\mathrm{Bias}\left(\bar{y}_{cr}\right) \approx \frac{1}{\bar{X}} \sum_{h=1}^{k} w_h^2 \gamma_h \bar{Y}_h^2 \left(\frac{R}{R_h^2} C_{xh}^2 - \rho_h \frac{C_{yh} C_{xh}}{R_h} \right) \tag{5}$$

$$\mathrm{MSE}\left(\bar{y}_{cr}\right) \approx \sum_{h=1}^{k} w_h^2 \gamma_h \bar{Y}_h^2 \left(C_{yh}^2 - 2\frac{R}{R_h} \rho_h C_{yh} C_{xh} + \frac{R^2}{R_h^2} C_{xh}^2 \right) \tag{6}$$

$$\mathrm{Bias}\left(\bar{y}_{sr}\right) \approx \sum_{h=1}^{k} w_h \gamma_h \bar{Y}_h \left(C_{xh}^2 - \rho_h C_{yh} C_{xh}\right) \tag{7}$$

$$\mathrm{MSE}\left(\bar{y}_{sr}\right) \approx \sum_{h=1}^{k} w_h^2 \gamma_h \bar{Y}_h^2 \left(C_{yh}^2 - 2\rho_h C_{yh} C_{xh} + C_{xh}^2\right) \tag{8}$$

เมื่อ C_{xh} เป็นสัมประสิทธิ์การแปรผันประชากรของตัวแปรช่วยในชั้นภูมิ h , ρ_h เป็นสัมประสิทธ์สหสัมพันธ์ระหว่างตัวแปรที่สนใจและตัวแปรช่วยในชั้นภูมิ h , $R_h = \bar{Y}_h / \bar{X}_h$ และ $R = \bar{Y} / \bar{X}$

Kadilar and Cingi (2003) ได้เสนอตัวประมาณค่าเฉลี่ยประชากรแบบอัตราส่วนในการเลือกตัวอย่างแบ่งชั้นภูมิอย่างง่าย เพื่อเพิ่มประสิทธิภาพของตัวประมาณโดยพัฒนาจากตัวประมาณในการเลือกตัวอย่างสุ่มแบบง่าย ของ Sisodia and Dwivedi (1981) ซึ่งเพิ่มสัมประสิทธิ์การแปรผันในตัวประมาณได้เป็น

$$\bar{y}_{cr.sd} = \bar{y}_{st} \frac{\sum_{h=1}^{k} w_h \left(\bar{X}_h + C_{xh}\right)}{\sum_{h=1}^{k} w_h \left(\bar{x}_h + C_{xh}\right)} \tag{9}$$

และตัวประมาณของ Upadphyaya and Singh (1999) ซึ่งเพิ่มสัมประสิทธิ์การแปรผันและสัมประสิทธิ์ความโด่งในตัวประมาณดังนี้

$$\bar{y}_{cr.us1} = \bar{y}_{st} \frac{\sum_{h=1}^{k} w_h \left(\bar{X}_h \beta_{2h}(x) + C_{xh}\right)}{\sum_{h=1}^{k} w_h \left(\bar{x}_h \beta_{2h}(x) + C_{xh}\right)} \tag{10}$$

$$\bar{y}_{cr.us2} = \bar{y}_{st} \frac{\sum_{h=1}^{k} w_h \left(\bar{X}_h C_{xh} + \beta_{2h}(x)\right)}{\sum_{h=1}^{k} w_h \left(\bar{x}_h C_{xh} + \beta_{2h}(x)\right)} \tag{11}$$

เมื่อ $\beta_{2h}(x)$ เป็นสัมประสิทธิ์ความโด่งประชากรของตัวแปรช่วยในชั้นภูมิ h
ความเอนเอียง และค่าคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณทั้งสาม เขียนได้ดังนี้

$$\mathrm{Bias}\left(\bar{y}_{cr.j}\right) \approx \frac{1}{\bar{X}_j} \sum_{h=1}^{k} w_h^2 \gamma_h \bar{Y}_h^2 \left(\frac{R_j}{R_h^2} C_{xh}^2 - \rho_h \frac{C_{yh} C_{xh}}{R_h}\right) \tag{12}$$

$$\mathrm{MSE}\left(\bar{y}_{cr.j}\right) \approx \sum_{h=1}^{k} w_h^2 \gamma_h \bar{Y}_h^2 \left(C_{yh}^2 - 2\frac{R_j}{R_h} \rho_h C_{yh} C_{xh} + \frac{R_j^2}{R_h^2} C_{xh}^2\right) \tag{13}$$

เมื่อ $R_j = \bar{y}_{st} / \bar{X}_j$, $j = sd, us1$ และ $us2$, $\bar{X}_{sd} = \sum_{h=1}^{k} w_h \left(\bar{X}_h + C_{xh} \right)$, $\bar{X}_{us1} = \sum_{h=1}^{k} w_h \left(\bar{X}_h \beta_{2h}(x) + C_{xh} \right)$
และ $\bar{X}_{us2} = \sum_{h=1}^{k} w_h \left(\bar{X}_h C_{xh} + \beta_{2h}(x) \right)$

ตัวประมาณ $\bar{y}_{cr.j}$, $j = sd, us1$ และ $us2$ มีประสิทธิภาพสูงกว่าตัวประมาณ $\bar{y}_{cr}$ เมื่อเป็นไปตามเงื่อนไขข้อใดข้อหนึ่งดังต่อไปนี้เท่านั้น

$$\sum_{h=1}^{k} w_h^2 \gamma_h C_{xh}^2 \bar{X}_h^2 < \frac{2\sum_{h=1}^{k} w_h^2 \gamma_h \rho_h C_{yh} C_{xh} \bar{Y}_h \bar{X}_h}{R_j + R} \; ; \; \sum_{h=1}^{k} w_h^2 \gamma_h C_{xh}^2 \bar{X}_h^2 > \frac{2\sum_{h=1}^{k} w_h^2 \gamma_h \rho_h C_{yh} C_{xh} \bar{Y}_h \bar{X}_h}{R_j + R}$$

ตัวประมาณ $\bar{y}_{cr.sd}$ ใน (9) เขียนขึ้นตามแนวทางของตัวประมาณแบบอัตราส่วนรวม แต่ถ้าเขียนขึ้นตามแนวทางของตัวประมาณแบบอัตราส่วนแยก จะได้ตัวประมาณแบบอัตราส่วนดังนี้

$$\bar{y}_{sr.sd} = \sum_{h=1}^{k} w_h \bar{y}_h \left(\frac{\bar{X}_h + C_{xh}}{\bar{x}_h + C_{xh}} \right) \tag{14}$$

ความเอนเอียง และค่าคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณ $\bar{y}_{sr.sd}$ เขียนได้ (Tailor and Lone, 2014) ดังนี้

$$\text{Bias}\left(\bar{y}_{sr.sd} \right) \approx \sum_{h=1}^{k} w_h \gamma_h \bar{Y}_h \left(\lambda_{1h}^2 C_{xh}^2 - \lambda_{1h} \rho_h C_{yh} C_{xh} \right) \tag{15}$$

$$\text{MSE}\left(\bar{y}_{sr.sd} \right) \approx \sum_{h=1}^{k} w_h^2 \gamma_h \bar{Y}_h^2 \left(C_{yh}^2 - 2\lambda_{1h} \rho_h C_{yh} C_{xh} + \lambda_{1h}^2 C_{xh}^2 \right) \tag{16}$$

เมื่อ $\lambda_{1h} = \bar{X}_h / \left(\bar{X}_h + C_{xh} \right)$

ต่อมา Kadilar and Cingi (2005) ได้เสนอตัวประมาณแบบอัตราส่วน $\bar{y}_{cr.kc}$ ที่มีประสิทธิภาพสูงกว่าตัวประมาณอัตราส่วนรวมเดิม $\bar{y}_{cr}$ ใน (3) ซึ่งนิยามดังนี้

$$\bar{y}_{cr.kc} = \kappa \frac{\bar{y}_{st}}{\bar{x}_{st}} \bar{X} \tag{17}$$

และ κ ที่ทำให้ค่าคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณ $\bar{y}_{cr.kc}$ มีค่าต่ำสุด จะเท่ากับ

$$\kappa = \frac{\bar{Y}^2}{\bar{Y}^2 + \sum_{h=1}^{k} w_h^2 \gamma_h \bar{Y}_h^2 \left(C_{yh}^2 - 2\frac{R}{R_h} \rho_h C_{yh} C_{xh} + \frac{R^2}{R_h^2} C_{xh}^2 \right)} \tag{18}$$

ความเอนเอียง และค่าคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณ $\bar{y}_{cr.kc}$ เขียนได้ดังนี้

$$\mathrm{Bias}\left(\bar{y}_{cr.kc}\right) \approx \frac{1}{\bar{X}} \sum_{h=1}^{k} w_h^2 \gamma_h \bar{Y}_h^2 \left(\frac{R}{R_h^2} C_{xh}^2 - \kappa \rho_h \frac{C_{yh} C_{xh}}{R_h} \right) + (\kappa - 1)\bar{Y} \tag{19}$$

$$\mathrm{MSE}\left(\bar{y}_{cr.kc}\right) \approx \kappa^2 \sum_{h=1}^{k} w_h^2 \gamma_h \bar{Y}_h^2 \left(C_{yh}^2 - 2\frac{R}{R_h} \rho_h C_{yh} C_{xh} + \frac{R^2}{R_h^2} C_{xh}^2 \right) + \left(\kappa - 1\right)^2 \bar{Y}^2 \tag{20}$$

อนึ่ง ถ้าตัวประมาณแบบอัตราส่วนไม่ได้มีรูปแบบดัง $\bar{y}_{cr.kc}$ ใน (17) อาจมีประสิทธิภาพสูงกว่าตัวประมาณ $\bar{y}_{cr.kc}$ ได้ เช่น ภายใต้เงื่อนไขบางประการ ตัวประมาณแบบอัตราส่วนที่ใช้ตัวประมาณของ Khoshnevisan *et al*. (2007) มีประสิทธิภาพสูงกว่า (Koyuncu and Kadilar, 2009) เป็นต้น

วิธีการดำเนินการวิจัย

จากการทบทวนวรรณกรรมเกี่ยวกับตัวประมาณในการเลือกตัวอย่างแบ่งชั้นภูมิอย่างง่ายดังกล่าวข้างต้น ยังไม่พบตัวประมาณที่พัฒนาจากตัวประมาณของซิลส์ (Searls, 1964)

$$\bar{y}_s = \frac{\bar{y}}{1+\gamma C_y^2} \tag{21}$$

ความเอนเอียง ความแปรปรวน และค่าคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณ $\bar{y}_s$ เท่ากับ

$$\mathrm{Bias}\left(\bar{y}_s\right) = -\frac{\gamma C_y^2 \bar{Y}}{1+\gamma C_y^2} \tag{22}$$

$$\mathrm{V}\left(\bar{y}_s\right) = \frac{\gamma C_y^2 \bar{Y}^2}{\left(1+\gamma C_y^2\right)^2} \tag{23}$$

$$\mathrm{MSE}\left(\bar{y}_s\right) = \frac{\gamma C_y^2 \bar{Y}^2}{1+\gamma C_y^2} \tag{24}$$

และตัวประมาณแบบอัตราส่วนที่พัฒนาด้วยวิธีของซิลส์ในการเลือกตัวอย่างสุ่มแบบง่ายที่ Jitthavech and Lorchirachoonkul (2013) นำเสนอไว้

$$\bar{y}_{rs} = \frac{1+\gamma C_x^2}{1+\gamma C_y^2}\frac{\bar{y}}{\bar{x}}\bar{X} \quad (25)$$

ความเอนเอียง ความแปรปรวน และค่าคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณ $\bar{y}_{rs}$ เขียนได้ดังนี้

$$\text{Bias}\left(\bar{y}_{rs}\right) \approx \frac{\gamma\bar{Y}}{1+\gamma C_y^2}\left[\left(1+\gamma C_x^2\right)\left(C_x^2-\rho C_y C_x\right)-\left(C_y^2-C_x^2\right)\right] \quad (26)$$

$$\text{MSE}\left(\bar{y}_{rs}\right) \approx \text{V}\left(\bar{y}_{rs}\right) \approx \left(\frac{1+\gamma C_x^2}{1+\gamma C_y^2}\right)^2 \text{MSE}\left(\bar{y}_r\right) \quad (27)$$

เมื่อ $\text{MSE}\left(\bar{y}_r\right) \approx \gamma\bar{Y}^2\left(C_y^2-2\rho C_y C_x+C_x^2\right)$ จึงกล่าวได้ว่าตัวประมาณ $\bar{y}_{rs}$ มีประสิทธิภาพสูงกว่าตัวประมาณ $\bar{y}_r$ เสมอ ถ้า $C_y > C_x$ ดังนั้น ด้วยเหตุผลทำนองเดียวกันที่นักวิจัยใช้ในการพัฒนาตัวประมาณในการเลือกตัวอย่างแบ่งชั้นภูมิ ดังที่ปรากฏในการทบทวนวรรณกรรม จึงเป็นที่น่าสนใจที่จะศึกษาตัวประมาณในการเลือกตัวอย่างแบ่งชั้นภูมิที่พัฒนาจากตัวประมาณของซิลส์ และตัวประมาณแบบอัตราส่วนที่พัฒนาด้วยวิธีของซิลส์ของ Jitthavech and Lorchirachoonkul

อนึ่ง ความเอนเอียง ความแปรปรวนและค่าคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณแบบอัตราส่วนข้างต้น พิสูจน์โดยการกระจายอนุกรมเทย์เลอร์ถึงอันดับที่หนึ่ง (First-Order Taylor Series Approximation) เท่านั้น ทำให้ความแปรปรวนและค่าคลาดคลื่อนกำลังสองเฉลี่ย มีค่าเท่ากัน เพื่อให้เห็นผลกระทบจากความเอนเอียงในการเปรียบเทียบประสิทธิภาพ การวิเคราะห์ในการศึกษานี้ จะกระทำโดยการกระจายอนุกรมเทย์เลอร์ถึงอันดับที่สอง แทนอันดับที่หนึ่ง ที่เป็นที่นิยมกัน ทฤษฎีที่จะนำเสนอในการศึกษาครั้งนี้ จะมี 4 บท ทฤษฎีบทที่ 1 เป็นการวิเคราะห์เกี่ยวกับตัวประมาณใหม่ซึ่งเป็นตัวประมาณในการเลือกตัวอย่างแบบแบ่งชั้นภูมิอย่างง่ายที่พัฒนาจากตัวประมาณของซิลส์ โดยไม่มีตัวแปรช่วย พร้อมบทตั้งที่ 1 ซึ่งเป็นเงื่อนไขที่ประสิทธิภาพของตัวประมาณใหม่สูงกว่าตัวประมาณค่าแบบดั้งเดิมที่ไม่มีตัวแปรช่วย ทฤษฎีบทที่ 2 และ 3 เป็นการวิเคราะห์เกี่ยวกับตัวประมาณแบบอัตราส่วนเดิมในการเลือกตัวอย่างสุ่มแบบง่ายและตัวประมาณค่าเฉลี่ยประชากรแบบอัตราส่วนแยกเดิมในการเลือกตัวอย่างแบบแบ่งชั้นภูมิอย่างง่ายตามลำดับ โดยการประมาณจะใช้การกระจายอนุกรมเทย์เลอร์ถึงอันดับที่สอง และทฤษฎีบทที่ 4 เป็นการวิเคราะห์เกี่ยวกับตัวประมาณใหม่ซึ่งเป็นตัวประมาณแบบอัตราส่วนที่ใช้ตัวประมาณค่าของซิลส์ พร้อมบทตั้งที่ 2 ซึ่งเป็นเงื่อนไขที่ประสิทธิภาพของตัวประมาณใหม่สูงกว่าตัวประมาณค่าเฉลี่ยประชากรแบบอัตราส่วนแยกเดิม

ทฤษฎีบทที่ 1 ในการเลือกตัวอย่างแบบแบ่งชั้นภูมิอย่างง่ายขนาด n จากประชากรขนาด N โดยไม่คืนที่ เมื่อทราบค่าสัมประสิทธิ์การแปรผันประชากรของตัวแปรที่สนใจในทุกชั้นภูมิ และทุกชั้นภูมิมีค่า $\gamma_h C_{yh}^2$ น้อยกว่า 1 จะได้ตัวประมาณค่าเฉลี่ยประชากรด้วยวิธีของซิลส์ เท่ากับ

$$\bar{y}_{st.s} = \sum_{h=1}^{k} w_h \frac{\bar{y}_h}{1+\gamma_h C_{yh}^2} \tag{28}$$

ความเอนเอียง ความแปรปรวน และค่าคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณ $\bar{y}_{st.s}$ เท่ากับ

$$\mathrm{Bias}\left(\bar{y}_{st.s}\right) = -\sum_{h=1}^{k} \frac{w_h \gamma_h C_{yh}^2 \bar{Y}_h}{1+\gamma_h C_{yh}^2} \tag{29}$$

$$\mathrm{V}\left(\bar{y}_{st.s}\right) = \sum_{h=1}^{k} \frac{w_h^2 \gamma_h C_{yh}^2 \bar{Y}_h^2}{\left(1+\gamma_h C_{yh}^2\right)^2} \tag{30}$$

$$\mathrm{MSE}\left(\bar{y}_{st.s}\right) = \sum_{h=1}^{k} \frac{w_h^2 \gamma_h C_{yh}^2 \bar{Y}_h^2}{\left(1+\gamma_h C_{yh}^2\right)^2} + \left[\sum_{h=1}^{k} \frac{w_h \gamma_h C_{yh}^2 \bar{Y}_h}{1+\gamma_h C_{yh}^2}\right]^2 \tag{31}$$

บทตั้งที่ 1 ตัวประมาณ $\bar{y}_{st.s}$ จะมีประสิทธิภาพสูงกว่าตัวประมาณ $\bar{y}_{st}$ ถ้า

$$\sum_{h=1}^{k} B_{sh}^2 \left(2+\gamma_h C_{yh}^2\right) > \left(\sum_{h=1}^{k} B_{sh}\right)^2 \tag{32}$$

เมื่อ $B_{sh} = \dfrac{w_h \gamma_h C_{yh}^2 \bar{Y}_h}{1+\gamma_h C_{yh}^2}$

ทฤษฎีบทที่ 2 ในการเลือกตัวอย่างสุ่มแบบง่ายขนาด n จากประชากรขนาด N โดยไม่คืนที่ และทราบสารสนเทศของตัวแปรช่วย จะได้ตัวประมาณค่าเฉลี่ยประชากรแบบอัตราส่วนเดิม เท่ากับ $\bar{y}_r = \dfrac{\bar{y}}{\bar{x}} \bar{X}$ (33)

ความเอนเอียง ความแปรปรวน และค่าคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณ $\bar{y}_r$ เท่ากับ

$$\mathrm{Bias}\left(\bar{y}_r\right) \approx \gamma \bar{Y}\left(C_x^2 - \rho C_y C_x\right) \tag{34}$$

$$\mathrm{V}\left(\bar{y}_r\right) \approx \bar{Y}^2 \left[\gamma\left(C_y^2 - 2\rho C_y C_x + C_x^2\right) - \gamma^2 \left(C_x^2 - \rho C_y C_x\right)^2 + o(n^{-2})\right] \tag{35}$$

$$\mathrm{MSE}\left(\bar{y}_r\right) \approx \bar{Y}^2\left[\gamma\left(C_y^2-2\rho C_y C_x+C_x^2\right)+o(n^{-2})\right] \tag{36}$$

เมื่อ $o(n^{-2})=\dfrac{\theta_1}{n^2}\left(4C_{12}-2C_{03}-2C_{21}\right)+\dfrac{\theta_2}{n^2}\left(3C_{02}C_{20}+6C_{11}^2+9C_{02}^2-18C_{02}C_{11}\right)$

โดยที่ $\theta_1=\dfrac{(N-n)(N-2n)}{(N-1)(N-2)}$, $\theta_2=\dfrac{N(N-n)(N-n-1)}{(N-1)(N-2)(N-3)}$ และ $C_{rs}=\dfrac{1}{N}\sum\limits_{i=1}^{N}\dfrac{\left(Y_i-\bar{Y}\right)^r\left(X_i-\bar{X}\right)^s}{\bar{Y}^r\bar{X}^s}$

ทฤษฎีบทที่ 3 ในการเลือกตัวอย่างแบบแบ่งชั้นภูมิอย่างง่ายขนาด n จากประชากรขนาด N โดยไม่คืนที่ และทราบสารสนเทศของตัวแปรช่วยในทุกชั้นภูมิ จะได้ตัวประมาณค่าเฉลี่ยประชากรแบบอัตราส่วนแยกเดิม เท่ากับ

$$\bar{y}_{sr}=\sum_{h=1}^{k} w_h \frac{\bar{y}_h}{\bar{x}_h}\bar{X}_h \tag{37}$$

ความเอนเอียง ความแปรปรวน และค่าคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณ $\bar{y}_{sr}$ จะเท่ากับ

$$\mathrm{Bias}\left(\bar{y}_{sr}\right) \approx \sum_{h=1}^{k} w_h\gamma_h\bar{Y}_h\left(C_{xh}^2-\rho_h C_{yh}C_{xh}\right) \tag{38}$$

$$\mathrm{V}\left(\bar{y}_{sr}\right) \approx \sum_{h=1}^{k} w_h^2\bar{Y}_h^2\left[\gamma_h\left(C_{yh}^2-2\rho_h C_{yh}C_{xh}+C_{xh}^2\right)-\gamma_h^2\left(C_{xh}^2-\rho_h C_{yh}C_{xh}\right)^2+o(n_h^{-2})\right] \tag{39}$$

$$\begin{aligned}\mathrm{MSE}\left(\bar{y}_{sr}\right) \approx{}& \sum_{h=1}^{k} w_h^2\bar{Y}_h^2\left[\gamma_h\left(C_{yh}^2-2\rho_h C_{yh}C_{xh}+C_{xh}^2\right)-\gamma_h^2\left(C_{xh}^2-\rho_h C_{yh}C_{xh}\right)^2+o(n_h^{-2})\right]\\ &+\left[\sum_{h=1}^{k} w_h\gamma_h\bar{Y}_h\left(C_{xh}^2-\rho_h C_{yh}C_{xh}\right)\right]^2\end{aligned} \tag{40}$$

เมื่อ $o(n_h^{-2})=\dfrac{\theta_{1h}}{n_h^2}\left(4C_{12h}-2C_{03h}-2C_{21h}\right)+\dfrac{\theta_{2h}}{n_h^2}\left(3C_{02h}C_{20h}+6C_{11h}^2+9C_{02h}^2-18C_{02h}C_{11h}\right)$

โดยที่ $\theta_{1h}=\dfrac{(N_h-n_h)(N_h-2n_h)}{(N_h-1)(N_h-2)}$, $\theta_{2h}=\dfrac{N_h(N_h-n_h)(N_h-n_h-1)}{(N_h-1)(N_h-2)(N_h-3)}$

และ $C_{rsh}=\dfrac{1}{N_h}\sum\limits_{i=1}^{N_h}\dfrac{\left(Y_{ih}-\bar{Y}_h\right)^r\left(X_{ih}-\bar{X}_h\right)^s}{\bar{Y}_h^r\bar{X}_h^s}$

ทฤษฎีบทที่ 4 ในการเลือกตัวอย่างแบบแบ่งชั้นภูมิอย่างง่ายขนาด n จากประชากรขนาด N โดยไม่คืนที่ เมื่อทราบค่าสัมประสิทธิ์การแปรผันประชากรของตัวแปรที่สนใจและตัวแปรช่วยในทุกชั้นภูมิ และทุกชั้นภูมิมีค่า $\gamma_h C_{yh}^2$ และ $\gamma_h C_{xh}^2$ น้อยกว่า 1 จะได้ตัวประมาณค่าเฉลี่ยประชากรแบบอัตราส่วนด้วยวิธีของซิลส์ เท่ากับ

$$\bar{y}_{sr.s} = \sum_{h=1}^{k} w_h \frac{1+\gamma_h C_{xh}^2}{1+\gamma_h C_{yh}^2} \frac{\bar{y}_h}{\bar{x}_h} \bar{X}_h \tag{41}$$

ความเอนเอียง ความแปรปรวน และค่าคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณ $\bar{y}_{sr.s}$ จะเท่ากับ

$$\mathrm{Bias}\left(\bar{y}_{sr.s}\right) \approx \sum_{h=1}^{k} w_h \frac{1+\gamma_h C_{xh}^2}{1+\gamma_h C_{yh}^2} \mathrm{Bias}\left(\bar{y}_{rh}\right) - \sum_{h=1}^{k} \frac{w_h \gamma_h \bar{Y}_h \left(C_{yh}^2 - C_{xh}^2\right)}{1+\gamma_h C_{yh}^2} \tag{42}$$

$$\mathrm{V}\left(\bar{y}_{sr.s}\right) \approx \sum_{h=1}^{k} w_h^2 \left(\frac{1+\gamma_h C_{xh}^2}{1+\gamma_h C_{yh}^2}\right)^2 \mathrm{V}\left(\bar{y}_{rh}\right) \tag{43}$$

$$\begin{aligned}\mathrm{MSE}\left(\bar{y}_{sr.s}\right) \approx{}& \sum_{h=1}^{k} w_h^2 \left(\frac{1+\gamma_h C_{xh}^2}{1+\gamma_h C_{xh}^2}\right)^2 \mathrm{V}\left(\bar{y}_{rh}\right) + \left[\sum_{h=1}^{k} w_h \frac{1+\gamma_h C_{xh}^2}{1+\gamma_h C_{yh}^2} \mathrm{Bias}\left(\bar{y}_{rh}\right)\right]^2 \\ & - \sum_{h=1}^{k} \frac{w_h \gamma_h \bar{Y}_h \left(C_{yh}^2 - C_{xh}^2\right)}{1+\gamma_h C_{yh}^2} \left\{\sum_{h=1}^{k} \frac{w_h \gamma_h \bar{Y}_h}{1+\gamma_h C_{yh}^2} \left[2\gamma_h C_{xh}^4 + 3C_{xh}^2 - C_{yh}^2 - 2\rho_h C_{yh} C_{xh} \left(1+\gamma_h C_{xh}^2\right)\right]\right\}\end{aligned} \tag{44}$$

เมื่อ $\mathrm{Bias}\left(\bar{y}_{rh}\right) \approx \gamma_h \bar{Y}_h \left(C_{xh}^2 - \rho_h C_{yh} C_{xh}\right)$

และ $\mathrm{V}\left(\bar{y}_{rh}\right) \approx \bar{Y}_h^2 \left[\gamma_h \left(C_{yh}^2 - 2\rho_h C_{yh} C_{xh} + C_{xh}^2\right) - \gamma_h^2 \left(C_{xh}^2 - \rho_h C_{yh} C_{xh}\right)^2 + o(n_h^{-2})\right]$

บทตั้งที่ 2 ตัวประมาณ $\bar{y}_{sr.s}$ จะมีประสิทธิภาพสูงกว่าตัวประมาณ $\bar{y}_{sr}$ ถ้าทุกชั้นภูมิมีค่า $C_{yh} > C_{xh}$ และ

$$\min_{1\le j\le k} w_j \gamma_j \bar{Y}_j \left(C_{yj}^2 - 2\rho_j C_{yj} C_{xj} + C_{xj}^2\right) > -\frac{1}{2} \sum_{h=1}^{k} w_h \gamma_h \bar{Y}_h \left(3C_{xh}^2 - 2\rho_h C_{yh} C_{xh} - C_{yh}^2\right) \tag{45}$$

ผลการวิจัยและวิจารณ์ผล

ตัวประมาณใหม่ $\bar{y}_{st.s}$ และ $\bar{y}_{sr.s}$ ในทฤษฎีบทที่ 1 และ 4 เป็นตัวประมาณที่ใช้สัมประสิทธิ์การแปรผันในตัวประมาณ ตัวประมาณใหม่ $\bar{y}_{st.s}$ ซึ่งไม่ใช้ตัวแปรช่วย จะประเมินประสิทธิภาพโดยเปรียบเทียบกับตัวประมาณดั้งเดิมที่ไม่ใช้ตัวแปรช่วย $\bar{y}_{st}$ เช่นกัน สำหรับตัวประมาณใหม่ $\bar{y}_{sr.s}$ จะประเมินประสิทธิภาพโดยเปรียบเทียบกับตัวประมาณแบบอัตราส่วนแยกเดิม

$\bar{y}_{sr}$ และตัวประมาณแบบอัตราส่วน $\bar{y}_{sr.sd}$ ซึ่งเป็นตัวประมาณอัตราส่วนแบบแยก และมีประสิทธิภาพสูงกว่า $\bar{y}_{cr.sd}$ ซึ่งเป็นตัวประมาณแบบอัตราส่วนรวม (Cochran, 1977) เหตุผลที่ไม่เปรียบเทียบกับตัวประมาณ $\bar{y}_{cr.us1}$ และ $\bar{y}_{cr.us2}$ เป็นเพราะตัวประมาณทั้งสองตัวนี้ ใช้ทั้งสัมประสิทธิ์การแปรผันและสัมประสิทธิ์ความโด่ง

การประเมินจะใช้ข้อมูลที่สร้างขึ้นจากการจำลองโดยใช้โปรแกรม R เวอร์ชัน 3.4.1 ที่มีประชากรย่อย 3 ชั้นภูมิ โดยชั้นภูมิที่ 1, 2 และ 3 มีขนาด 20,000, 30,000 และ 50,000 หน่วยตามลำดับ ซึ่งแต่ละชั้นภูมิมี Y เป็นตัวแปรที่สนใจ และ X เป็นตัวแปรช่วย ที่มีความสัมพันธ์กัน โดยกำหนดค่าสัมประสิทธิ์สหสัมพันธ์ในแต่ละชั้นภูมิไว้ 3 ค่า คือ 0.6, 0.7 และ 0.8 และกำหนดให้ตัวแปรทั้ง 2 ตัว มีการแจกแจงแกมมา เพื่อให้ความแตกต่างของประสิทธิภาพของตัวประมาณได้อย่างชัดเจน โดย $Y_1 \sim Gamma(\alpha = 0.5, \beta = 575)$, $X_1 \sim Gamma(\alpha = 0.6, \beta = 230)$, $Y_2 \sim Gamma(\alpha = 0.25, \beta = 2850)$, $X_2 \sim Gamma(\alpha = 0.3, \beta = 1200)$, $Y_3 \sim Gamma(\alpha = 0.4, \beta = 190)$ และ $X_3 \sim Gamma(\alpha = 0.49, \beta = 35)$ ค่าสถิติต่างๆ ของประชากรในชั้นภูมิ ที่เกี่ยวข้องกับการศึกษา ได้แสดงไว้ในตารางที่ 1

ตารางที่ 1 ค่าพารามิเตอร์ จำแนกตามค่าสัมประสิทธิ์สหสัมพันธ์และชั้นภูมิ

ρ_h	ชั้นภูมิ	N_h	$\bar{Y}_h$	$\bar{X}_h$	C_{yh}	C_{xh}	ρ_h
0.6	1	20,000	282.95	137.47	1.40	1.29	0.58
	2	30,000	695.01	354.47	1.97	1.82	0.58
	3	50,000	74.92	17.07	1.58	1.43	0.59
	รวม	100,000	302.55	142.37	2.72	2.74	0.63
0.7	1	20,000	283.59	137.16	1.40	1.30	0.68
	2	30,000	696.23	353.82	1.98	1.82	0.68
	3	50,000	75.23	17.08	1.56	1.42	0.69
	รวม	100,000	303.20	142.12	2.72	2.74	0.72
0.8	1	20,000	282.83	136.54	1.40	1.29	0.79
	2	30,000	697.59	354.54	1.97	1.83	0.79
	3	50,000	75.21	17.05	1.57	1.42	0.80
	รวม	100,000	303.45	142.19	2.71	2.76	0.81

การจำลอง จะใช้จำนวน 50 ตัวอย่างใน 3 ขนาดตัวอย่างคือ 100, 150 และ 200 หน่วยซึ่งจะสุ่มแบบง่ายโดยไม่คืนที่จากประชากรทั้ง 3 ชั้นภูมิโดยจัดสรรขนาดตัวอย่างในแต่ละชั้นภูมิตามขนาดของชั้นภูมิ ตัวอย่างที่ใช้ในการศึกษาเป็นตัวอย่างที่สอดคล้องกับเงื่อนไขตามบทตั้งที่ 1 และ 2 แล้วแต่กรณีเพื่อยืนยันความถูกต้องของการวิเคราะห์ การศึกษา จะกระทำทั้งในกรณีทราบและไม่ทราบค่าสัมประสิทธิ์การแปรผันประชากร ผลการจำลองที่สรุปจาก 50 ตัวอย่างได้แสดงไว้ตารางที่ 2-6 ตารางที่ 2 เป็นการเปรียบเทียบตัวประมาณ $\bar{y}_{st.s}$ กับ $\bar{y}_{st}$ ทั้งในกรณีที่ทราบค่าและไม่ทราบค่าสัมประสิทธิ์การแปรผันประชากร พบว่าตัวประมาณ $\bar{y}_{st.s}$ และ $\bar{y}_{st}$ มีค่า MSE ลดลงเมื่อขนาดตัวอย่างเพิ่มจาก 100 เป็น 200 ประสิทธิภาพของ $\bar{y}_{st.s}$ สัมพัทธ์กับ $\bar{y}_{st}$ มีค่ามาก 100 ในทั้ง 3 ขนาดตัวอย่างทั้งสองกรณีที่ศึกษา จึงสรุปได้ว่า $\bar{y}_{st.s}$ มีประสิทธิภาพสูงกว่า $\bar{y}_{st}$ ภายใต้เงื่อนไขในบทตั้งที่ 1 ผลการเปรียบเทียบตัวประมาณแบบอัตราส่วนที่เสนอใหม่ $\bar{y}_{sr.s}$ กับตัวประมาณแบบอัตราส่วนแยกเดิม $\bar{y}_{sr}$ เมื่อทราบและไม่ทราบค่าสัมประสิทธิ์การแปรผันประชากร ได้แสดงไว้ในตารางที่ 3 และ 4 ตามลำดับ สรุปได้ใน

ทำนองเดียวกันกับข้างต้น ประสิทธิภาพของ $\bar{y}_{sr.s}$ สัมพัทธ์กับ $\bar{y}_{sr}$ มีค่ามากกว่า 100 ในทั้ง 3 ขนาดตัวอย่างและ 3 ค่าของสัมประสิทธิ์สหสัมพันธ์ ทั้งกรณีที่ทราบค่าและไม่ทราบค่าสัมประสิทธิ์การแปรผันประชากร จึงสรุปได้ว่า $\bar{y}_{sr.s}$ มีประสิทธิภาพสูงกว่า $\bar{y}_{sr}$ ภายใต้เงื่อนไขในบทตั้งที่ 2 ในการเปรียบเทียบประสิทธิภาพระหว่างตัวประมาณ $\bar{y}_{sr.s}$ กับ $\bar{y}_{sr.sd}$ จะใช้การประมาณถึงอันดับที่หนึ่งเท่านั้น ทั้งนี้เพราะการวิเคราะห์เดิม ใช้การประมาณถึงอันดับที่หนึ่ง (Kadilar and Cingi, 2003; Tailor and Lone, 2014) ผลการจำลองได้แสดงไว้ในตารางที่ 5 และ 6 ซึ่งสรุปได้ทำนองเดียวกันกับการเปรียบเทียบทั้งสองข้างต้น ประสิทธิภาพของ $\bar{y}_{sr.s}$ สัมพัทธ์กับ $\bar{y}_{sr.sd}$ มีค่ามากกว่า 100 ในทั้ง 3 ขนาดตัวอย่าง 3 ค่าของสัมประสิทธิ์สหสัมพันธ์ ทั้งกรณีที่ทราบค่าและไม่ทราบค่าสัมประสิทธิ์การแปรผันประชากร จึงสรุปได้ว่า $\bar{y}_{sr.s}$ มีประสิทธิภาพสูงกว่า $\bar{y}_{sr.sd}$ ในการจำลอง นอกจากนี้ ความเอนเอียงของตัวประมาณ $\bar{y}_{sr.s}$ ยังต่ำกว่าความเอนเอียงของ $\bar{y}_{sr.sd}$ อีกด้วย

***ตารางที่** 2* ค่าประมาณค่าเฉลี่ยประชากร ความเอนเอียง ค่าคลาดเคลื่อนกำลังสองเฉลี่ย และค่าประสิทธิภาพสัมพัทธ์ของตัวประมาณเทียบกับ $\bar{y}_{st}$ จำแนกตามตัวประมาณค่า ขนาดตัวอย่าง เมื่อทราบและไม่ทราบสัมประสิทธิ์การแปรผันประชากร

สัมประสิทธิ์การแปรผัน	**ขนาดตัวอย่าง**	**ตัวประมาณ**	**ค่าประมาณ**	**ความเอนเอียง**	MSE	**ประสิทธิภาพสัมพัทธ์ (%)**
ทราบ	100	$\bar{y}_{st}$	287.64	0.00	6,094.78	100.00
		$\bar{y}_{st.s}$	258.23	-29.41	5,708.52	106.77
	150	$\bar{y}_{st}$	303.70	0.00	4,365.81	100.00
		$\bar{y}_{st.s}$	282.32	-21.38	4,188.06	104.24
	200	$\bar{y}_{st}$	317.60	0.00	3,554.71	100.00
		$\bar{y}_{st.s}$	300.46	-17.14	3,444.01	103.21
ไม่ทราบ	100	$\bar{y}_{st}$	287.64	0.00	4,869.54	100.00
		$\bar{y}_{st.s}$	263.17	-24.47	4,626.92	105.24
	150	$\bar{y}_{st}$	303.70	0.00	3,795.18	100.00
		$\bar{y}_{st.s}$	284.77	-18.93	3,655.41	103.82
	200	$\bar{y}_{st}$	317.60	0.00	3,178.77	100.00
		$\bar{y}_{st.s}$	302.02	-15.58	3,086.44	102.99

ตารางที่ 3 ค่าประมาณค่าเฉลี่ยประชากร ความเอนเอียง ค่าคลาดเคลื่อนกำลังสองเฉลี่ย และประสิทธิภาพสัมพัทธ์ของตัวประมาณเทียบกับ $\bar{y}_{sr}$ จำแนกตามตัวประมาณค่า ขนาดตัวอย่าง สัมประสิทธิ์สหสัมพันธ์ เมื่อทราบสัมประสิทธิ์การแปรผันประชากร

สัมประสิทธิ์สหสัมพันธ์	ขนาดตัวอย่าง	ค่าประมาณ		ความเอนเอียง		MSE		ประสิทธิภาพสัมพัทธ์ (%)
		$\bar{y}_{sr}$	$\bar{y}_{sr.s}$	$\bar{y}_{sr}$	$\bar{y}_{sr.s}$	$\bar{y}_{sr}$	$\bar{y}_{sr.s}$	
0.6	100	290.71	286.21	9.27	4.65	5,606.33	5,351.05	104.77
	150	311.26	307.93	6.66	3.33	3,809.16	3,685.66	103.35
	200	306.08	303.57	5.19	2.53	2,961.98	2,887.82	102.57
0.7	100	317.49	312.46	7.17	2.18	4,946.56	4,727.04	104.64
	150	309.55	306.16	4.89	1.48	3,069.89	2,973.13	103.25
	200	303.98	301.45	3.53	0.94	2,083.23	2,032.82	102.48
0.8	100	315.15	310.49	4.03	-0.76	3,358.37	3,236.15	103.78
	150	310.15	306.98	2.84	-0.43	2,148.18	2,092.69	102.65
	200	307.58	305.18	1.89	-0.45	1,252.38	1,227.89	101.99

ตารางที่ 4 ค่าประมาณค่าเฉลี่ยประชากร ความเอนเอียง ค่าคลาดเคลื่อนกำลังสองเฉลี่ย และประสิทธิภาพสัมพัทธ์ของตัวประมาณเทียบกับ $\bar{y}_{sr}$ จำแนกตามตัวประมาณค่า ขนาดตัวอย่าง สัมประสิทธิ์สหสัมพันธ์ เมื่อไม่ทราบสัมประสิทธิ์การแปรผันประชากร

สัมประสิทธิ์สหสัมพันธ์	ขนาดตัวอย่าง	ค่าประมาณ		ความเอนเอียง		MSE		ประสิทธิภาพสัมพัทธ์ (%)
		$\bar{y}_{sr}$	$\bar{y}_{sr.s}$	$\bar{y}_{sr}$	$\bar{y}_{sr.s}$	$\bar{y}_{sr}$	$\bar{y}_{sr.s}$	
0.6	100	290.71	285.17	6.33	0.64	3,888.04	3,702.66	105.01
	150	311.26	306.87	5.01	0.82	3,065.66	2,960.45	103.55
	200	306.08	302.98	4.09	0.78	2,489.73	2,422.55	102.77
0.7	100	317.49	312.86	5.08	0.49	3,302.23	3,185.35	103.67
	150	309.55	306.38	3.96	0.74	2,475.67	2,407.96	102.81
	200	303.98	300.91	2.80	-0.29	1,844.22	1,793.20	102.85
0.8	100	315.15	311.82	3.46	0.08	2,441.44	2,380.38	102.57
	150	310.15	307.52	2.57	-0.13	1,841.09	1,803.16	102.10
	200	307.58	305.74	1.80	0.04	1,051.88	1,036.34	101.50

ตารางที่ 5 ค่าประมาณค่าเฉลี่ยประชากร ความเอนเอียง ค่าคลาดเคลื่อนกำลังสองเฉลี่ย และประสิทธิภาพสัมพัทธ์ของตัวประมาณเทียบกับ $\bar{y}_{sr.sd}$ จำแนกตามตัวประมาณค่า ขนาดตัวอย่าง สัมประสิทธิ์สหสัมพันธ์ เมื่อทราบสัมประสิทธิ์การแปรผันประชากร

สัมประสิทธิ์สหสัมพันธ์	**ขนาดตัวอย่าง**	**ค่าประมาณ**		**ความเอนเอียง**		MSE		**ประสิทธิภาพสัมพัทธ์ (%)**
		$\bar{y}_{sr.sd}$	$\bar{y}_{sr.s}$	$\bar{y}_{sr.sd}$	$\bar{y}_{sr.s}$	$\bar{y}_{sr.sd}$	$\bar{y}_{sr.s}$	
0.6	100	290.26	286.21	8.93	4.65	4,287.78	4,161.52	103.03
	150	310.92	307.93	6.42	3.33	3,150.21	3,090.58	101.93
	200	306.05	303.57	5.00	2.53	2,553.48	2,519.05	101.37
0.7	100	317.34	312.46	6.83	2.18	3,797.93	3,681.20	103.17
	150	309.38	306.16	4.67	1.48	2,584.08	2,532.05	102.05
	200	303.78	301.45	3.36	0.94	1,842.10	1,815.51	101.46
0.8	100	315.01	310.49	3.71	-0.76	2,658.93	2,583.45	102.92
	150	310.06	306.98	2.63	-0.43	1,849.30	1,814.93	101.89
	200	307.34	305.18	1.74	-0.45	1,135.52	1,120.22	101.37

หมายเหตุ: ค่า MSE ที่ใช้การกระจายอนุกรมเทย์เลอร์ถึงอันดับที่หนึ่งในการพิสูจน์

ตารางที่ 6 ค่าประมาณค่าเฉลี่ยประชากร ความเอนเอียง ค่าคลาดเคลื่อนกำลังสองเฉลี่ย และประสิทธิภาพสัมพัทธ์ของตัวประมาณเทียบกับ $\bar{y}_{sr.sd}$ จำแนกตามตัวประมาณค่า ขนาดตัวอย่าง สัมประสิทธิ์สหสัมพันธ์ เมื่อไม่ทราบสัมประสิทธิ์การแปรผันประชากร

สัมประสิทธิ์สหสัมพันธ์	**ขนาดตัวอย่าง**	**ค่าประมาณ**		**ความเอนเอียง**		MSE		**ประสิทธิภาพสัมพัทธ์ (%)**
		$\bar{y}_{sr.sd}$	$\bar{y}_{sr.s}$	$\bar{y}_{sr.sd}$	$\bar{y}_{sr.s}$	$\bar{y}_{sr.sd}$	$\bar{y}_{sr.s}$	
0.6	100	290.31	285.17	6.10	0.64	3,368.63	3,240.32	103.96
	150	310.93	306.87	4.83	0.82	2,692.38	2,626.25	102.52
	200	306.04	302.98	3.93	0.78	2,234.44	2,193.14	101.88
0.7	100	317.35	312.86	4.86	0.49	2,834.03	2,760.50	102.66
	150	309.40	306.38	3.78	0.74	2,176.36	2,135.43	101.92
	200	303.78	300.91	2.66	-0.29	1,666.92	1,633.18	102.07
0.8	100	315.01	311.82	3.22	0.08	2,081.73	2,042.81	101.91
	150	310.06	307.52	2.39	-0.13	1,627.50	1,605.72	101.36
	200	307.35	305.74	1.67	0.04	978.53	970.03	100.88

หมายเหตุ: ค่า MSE ที่ใช้การกระจายอนุกรมเทย์เลอร์ถึงอันดับที่หนึ่งในการพิสูจน์

สรุปผลการวิจัย

การศึกษานี้ได้เสนอตัวประมาณใหม่ 2 ตัวเพื่อใช้ในการเลือกตัวอย่างแบบแบ่งชั้นภูมิอย่างง่าย ตัวประมาณตัวที่หนึ่ง $\bar{y}_{st.s}$ เป็น ตัวประมาณที่ไม่ใช้ตัวแปรช่วยพัฒนาจากตัวประมาณของซิลส์ และตัวประมาณตัวที่สอง $\bar{y}_{sr.s}$ เป็นตัวประมาณแบบอัตราส่วนพัฒนาจากตัวประมาณแบบอัตราส่วนด้วยวิธีของซิลส์ของ Jitthavech and Lorchirachoonkul พร้อมด้วยคุณสมบัติของตัวประมาณค่าและเงื่อนไขที่มีประสิทธิภาพสูงกว่าตัวประมาณดั้งเดิมในกรณีที่ไม่ใช้ตัวแปรช่วยและตัวประมาณแบบอัตราส่วนแยกเดิม ผลการจำลอง ยืนยันความสอดคล้องของเงื่อนไขที่ได้วิเคราะห์ในบทตั้งที่ 1 และ 2 นอกจากนี้ ยังเปรียบเทียบประสิทธิภาพของตัวประมาณ $\bar{y}_{sr.s}$ กับตัวประมาณแบบอัตราส่วน $\bar{y}_{sr.sd}$ ที่ปรากฏในวรรณกรรมวิชาการชั้นนำ ซึ่งผลการจำลองสรุปได้ว่าตัวประมาณ $\bar{y}_{sr.s}$ มีประสิทธิภาพสูงกว่าตัวประมาณ $\bar{y}_{sr.sd}$ และความเอนเอียงยังต่ำกว่าอีกด้วย

การศึกษาตัวประมาณเพื่อใช้ในการเลือกตัวอย่างแบบแบ่งชั้นภูมิอย่างง่าย ควรพิจารณาการพัฒนาจากตัวประมาณที่มีประสิทธิภาพในการเลือกตัวอย่างสุ่มแบบง่ายที่มีมากมายในปัจจุบัน โดยเฉพาะตัวประมาณแบบอัตราส่วนที่มีรูปแบบหลากหลายแตกต่างไปจากรูปแบบตัวประมาณแบบอัตราส่วนเดิม ทั้งนี้เพื่อเพิ่มประสิทธิภาพของตัวประมาณในการเลือกตัวอย่างแบบแบ่งชั้นภูมิอย่างง่าย

เอกสารอ้างอิง

Cochran, W.G. (1977). *Sampling Tehniques*. (3rd ed). New York: John Wiley and Sons.

Jitthavech, J., & Lorchirachoonkul, V. (2013). Estimators in simple random sampling: Searls approach. *Songklanakarin Journal of Science and Technology*, *35*(6), 749-760.

Kadilar, C., & Cingi, H. (2003). Ratio estimators in stratified random sampling. *Biometrical Journal*, *45*(2), 218-225.

Kadilar, C., & Cingi, H. (2005). A New ratio estimator in stratified random sampling. *Communication in Statistics-Theory and Method*, *34*(3), 597-602.

Khoshnevisan, M., Singh, R., Chauhan, P., Sawan, N., & Smarandache, F. (2007). A general family of estimators for estimating population mean using known value of some population parameter(s). *Far East Journal of Theoretical Statistics*, *22*(2), 181-191.

Koyuncu, N., & Kadilar, C. (2009). Ratio and product estimators in stratified random sampling. *Journal of Statistical Planning and Inference*, *139*(1), 2552-2558.

Sangngam, P. (2014). Ratio estimators using coefficient of varation and coefficient of correlation. *Modern Applied Science*, *8*(5), 70-79.

Searls, D.T. (1964). The utilization of a known coefficient of variation in the estimation procedure. *American Statistical Association Journal*, *59*(308), 1225-1226.

Sharma, P., Verma, H.K., Sanaullah, A., & Singh, R. (2013). Some exponential ratio-product type estimators using information on auxiliary attributes under second order approximation. *International Journal of Statistics and Economics*, *12*(3), 58-66.

Sisodia, B.V.S., & Dwivedi, V.K. (1981). A modified ratio estimator using coefficient of variation of auxiliary variable. *Journal of the Indian Society of Agricultural Statistics*, *33*(1), 13-18.

Tailor, R., & Lone, H.A. (2014). Separate ratio-type estimators of populaton mean is stratified random sampling. Journal of Modern Applied Statistical Methods, *13*(1), 223-233

Upadhyaya, L.N., & Singh, H.P. (1999). Use of transformed auxiliary variable in estimating the finite population mean. *Biometrical Journal*, *41*(5), 627-636.

ภาคผนวก

บทพิสูจน์ทฤษฎีบทที่ 1

ความเอนเอียงของตัวประมาณ $\bar{y}_{st.s}$ สามารถเขียนได้จากคำนิยามของ $\bar{y}_{st.s}$ ใน (28) และ $\mathrm{Bias}\left(\bar{y}_s\right)$ ใน (22) เป็น

$$\mathrm{Bias}\left(\bar{y}_{st.s}\right)=\sum_{h=1}^{k} w_h \mathrm{Bias}\left(\frac{\bar{y}_h}{1+\gamma_h C_{yh}^2}\right)=-\sum_{h=1}^{k}\frac{w_h\gamma_h C_{yh}^2\bar{Y}_h}{1+\gamma_h C_{yh}^2}$$

ความแปรปรวนของตัวประมาณ $\bar{y}_{st.s}$ สามารถเขียนได้จาก (28) เช่นกันและ $\mathrm{V}\left(\bar{y}_s\right)$ ใน (23) เป็น

$$\mathrm{V}\left(\bar{y}_{st.s}\right)=\sum_{h=1}^{k} w_h^2 \mathrm{V}\left(\frac{\bar{y}_h}{1+\gamma_h C_{yh}^2}\right)=\sum_{h=1}^{k}\frac{w_h^2\gamma_h C_{yh}^2\bar{Y}_h^2}{\left(1+\gamma_h C_{yh}^2\right)^2}$$

ดังนั้น ค่าคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณ $\bar{y}_{st.s}$ จึงเท่ากับ

$$\mathrm{MSE}\left(\bar{y}_{st.s}\right)=\sum_{h=1}^{k}\frac{w_h^2\gamma_h C_{yh}^2\bar{Y}_h^2}{\left(1+\gamma_h C_{yh}^2\right)^2}+\left[\sum_{h=1}^{k}\frac{w_h\gamma_h C_{yh}^2\bar{Y}_h}{1+\gamma_h C_{yh}^2}\right]^2 \qquad \square$$

บทพิสูจน์บทตั้งที่ 1

ตัวประมาณ $\bar{y}_{st.s}$ จะมีประสิทธิภาพสูงกว่าตัวประมาณ $\bar{y}_{st}$ เมื่อ $\mathrm{MSE}\left(\bar{y}_{st}\right)>\mathrm{MSE}\left(\bar{y}_{st.s}\right)$

$$\sum_{h=1}^{k} w_h^2\gamma_h C_{yh}^2\bar{Y}_h^2>\sum_{h=1}^{k}\frac{w_h^2\gamma_h C_{yh}^2\bar{Y}_h^2}{\left(1+\gamma_h C_{yh}^2\right)^2}+\left[\sum_{h=1}^{k}\frac{w_h\gamma_h C_{yh}^2\bar{Y}_h}{1+\gamma_h C_{yh}^2}\right]^2$$

$$\sum_{h=1}^{k} w_h^2\gamma_h C_{yh}^2\bar{Y}_h^2\left[1-\frac{1}{\left(1+\gamma_h C_{yh}^2\right)^2}\right]>\left[\sum_{h=1}^{k}\frac{w_h\gamma_h C_{yh}^2\bar{Y}_h}{1+\gamma_h C_{yh}^2}\right]^2$$

$$\sum_{h=1}^{k} B_{sh}^2\left(2+\gamma_h C_{yh}^2\right)>\left(\sum_{h=1}^{k} B_{sh}\right)^2 \quad \text{เมื่อ} \quad B_{sh}=\frac{w_h\gamma_h C_{yh}^2\bar{Y}_h}{1+\gamma_h C_{yh}^2} \qquad \square$$

บทพิสูจน์ทฤษฎีบทที่ 2

ตัวประมาณค่าเฉลี่ยประชากรแบบอัตราส่วนเดิมในเลือกตัวอย่างสุ่มแบบง่าย เท่ากับ $\bar{y}_r=\frac{\bar{y}}{\bar{x}}\bar{X}$ (ก.1)

กำหนดให้ $\delta_0=\frac{\bar{y}-\bar{Y}}{\bar{Y}}$, $\delta_1=\frac{\bar{x}-\bar{X}}{\bar{X}}$ จะได้ค่าคาดหมาย (Sharma et la., 2013) ดังนี้

$$\mathrm{E}\left(\delta_0\right)=0,\ \mathrm{E}\left(\delta_1\right)=0,\ \mathrm{E}\left(\delta_0^2\right)=\frac{\theta_0}{n}C_{20},\ \mathrm{E}\left(\delta_1^2\right)=\frac{\theta_0}{n}C_{02},\ \mathrm{E}\left(\delta_0\delta_1\right)=\frac{\theta_0}{n}C_{11},$$

$$\mathrm{E}\left(\delta_0^2\delta_1\right)=\frac{\theta_1}{n^2}C_{21},\ \mathrm{E}\left(\delta_0\delta_1^2\right)=\frac{\theta_1}{n^2}C_{12},\ \mathrm{E}\left(\delta_1^3\right)=\frac{\theta_1}{n^2}C_{03},\ \mathrm{E}\left(\delta_0^2\delta_1^2\right)=\frac{\theta_2}{n^2}\left(C_{02}C_{20}+2C_{11}^2\right),$$

$$\mathrm{E}\left(\delta_0\delta_1^3\right)=3\frac{\theta_2}{n^2}C_{02}C_{11} \text{ และ } \mathrm{E}\left(\delta_1^4\right)=3\frac{\theta_2}{n^2}C_{02}^2 \quad \text{เมื่อ} \quad \theta_0=\frac{N-n}{N-1}$$

ตัวประมาณ $\bar{y}_r$ ใน (ก.1) เมื่อใช้สัญลักษณ์ δ_0 และ δ_1 จะเขียนได้ดังนี้

$$\bar{y}_r=\bar{Y}\left(1+\delta_0\right)\left(1+\delta_1\right)^{-1}$$

กระจายอนุกรมเทย์เลอร์ของ $\left(1+\delta_1\right)^{-1}$ จะได้

$$\begin{aligned}\bar{y}_r &= \bar{Y}\left(1+\delta_0\right)\left(1-\delta_1+\delta_1^2-\delta_1^3+\ldots\right)\\ &= \bar{Y}\left[1+\left(\delta_0-\delta_1\right)-\delta_1\left(\delta_0-\delta_1\right)+\delta_1^2\left(\delta_0-\delta_1\right)-\delta_1^3\left(\delta_0-\delta_1\right)+\ldots\right] \qquad (ก.2)\end{aligned}$$

เมื่อเก็บเฉพาะพจน์ของ δ_0 และ δ_1 ไม่เกินองศาที่สี่ ค่าคาดหมายของ $\bar{y}_r$ เมื่อประมาณถึงอันดับที่สอง จะได้

$$\begin{aligned}\mathrm{E}\left(\bar{y}_r\right) &\approx \bar{Y}\mathrm{E}\left[1+\left(\delta_0-\delta_1\right)-\delta_1\left(\delta_0-\delta_1\right)+\delta_1^2\left(\delta_0-\delta_1\right)-\delta_1^3\left(\delta_0-\delta_1\right)\right]\\ &\approx \bar{Y}\left[1+\frac{\theta_0}{n}\left(C_{02}-C_{11}\right)+\frac{\theta_1}{n^2}\left(C_{12}-C_{03}\right)+\frac{\theta_2}{n^2}\left(3C_{02}^2-3C_{02}C_{11}\right)\right] \qquad (ก.3)\end{aligned}$$

ความเอนเอียงของตัวประมาณ $\bar{y}_r$ เขียนได้เท่ากับ

$$\mathrm{Bias}\left(\bar{y}_r\right) \approx \bar{Y}\left[\frac{\theta_0}{n}\left(C_{02}-C_{11}\right)+\frac{\theta_1}{n^2}\left(C_{12}-C_{03}\right)+\frac{\theta_2}{n^2}\left(3C_{02}^2-3C_{02}C_{11}\right)\right] \qquad (ก.4)$$

พจน์ $\frac{\theta_0}{n}\left(C_{02}-C_{11}\right)$ สามารถพิสูจน์ได้โดยง่าย เท่ากับ $\frac{\theta_0}{n}\left(C_{02}-C_{11}\right)=\gamma\left(C_x^2-\rho C_y C_x\right)$

ซึ่งเมื่อแทนกลับไปใน (ก.4) จะสามารถเขียน ความเอนเอียงของตัวประมาณ $\bar{y}_r$ เป็นผลบวกของความเอนเอียงของตัวประมาณ $\bar{y}_r$ เมื่อประมาณถึงอันดับที่หนึ่ง และพจน์ในอันดับที่สอง ดังนี้

$$\mathrm{Bias}\left(\bar{y}_r\right) \approx \bar{Y}\left[\gamma\left(C_x^2-\rho C_y C_x\right)+\frac{\theta_1}{n^2}\left(C_{12}-C_{03}\right)+\frac{\theta_2}{n^2}\left(3C_{02}^2-3C_{02}C_{11}\right)\right]$$

ยกกำลังสองตัวประมาณ $\bar{y}_r$ ใน (ก.2) แล้วเก็บเฉพาะพจน์ของ δ_0 และ δ_1 ไม่เกินองศาที่สี่ และใส่ค่าคาดหมาย จะได้

$$\begin{aligned}\mathrm{E}\left(\bar{y}_r^2\right) &\approx \bar{Y}^2\mathrm{E}\Big[1+\left(\delta_0-\delta_1\right)^2+\delta_1^2\left(\delta_0-\delta_1\right)^2+2\left(\delta_0-\delta_1\right)-2\delta_1\left(\delta_0-\delta_1\right)+2\delta_1^2\left(\delta_0-\delta_1\right)\\ &\qquad -2\delta_1^3\left(\delta_0-\delta_1\right)-2\delta_1\left(\delta_0-\delta_1\right)^2+2\delta_1^2\left(\delta_0-\delta_1\right)^2\Big]\\ &\approx \bar{Y}^2\Big[1+\frac{\theta_0}{n}\left(C_{20}-4C_{11}+3C_{02}\right)+\frac{\theta_1}{n^2}\left(-2C_{21}+6C_{12}-4C_{03}\right)\\ &\qquad +\frac{\theta_2}{n^2}\left(3C_{02}C_{20}+6C_{11}^2-24C_{02}C_{11}+15C_{02}^2\right)\Big] \qquad (ก.5)\end{aligned}$$

ค่าคาดหมายของตัวประมาณ $\bar{y}_r$ ยกกำลังสอง สามารถเขียนได้จาก (ก.3) เท่ากับ

$$\mathrm{E}^2\left(\bar{y}_r\right) \approx \bar{Y}^2\left[1+\frac{\theta_0^2}{n^2}\left(C_{02}-C_{11}\right)^2+\frac{\theta_0}{n}\left(2C_{02}-2C_{11}\right)+\frac{\theta_1}{n^2}\left(2C_{12}-2C_{03}\right)+\frac{\theta_2}{n^2}\left(6C_{02}^2-6C_{02}C_{11}\right)\right] \qquad (ก.6)$$

ความแปรปรวนของตัวประมาณ $\bar{y}_r$ สามารถเขียนได้จาก (ก.5) และ (ก.6) เท่ากับ

$$\mathrm{V}\left(\bar{y}_r\right) \approx \bar{Y}^2\left[\gamma\left(C_y^2-2\rho C_y C_x+C_x^2\right)-\gamma^2\left(C_x^2-\rho C_y C_x\right)^2+o\left(n^{-2}\right)\right]$$

เมื่อ $o(n^{-2})=\frac{\theta_1}{n^2}\left(4C_{12}-2C_{03}-2C_{21}\right)+\frac{\theta_2}{n^2}\left(3C_{02}C_{20}+6C_{11}^2+9C_{02}^2-18C_{02}C_{11}\right)$

ดังนั้น ค่าคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณ $\bar{y}_r$ จึงเท่ากับ

$$\mathrm{MSE}\left(\bar{y}_r\right) \approx \bar{Y}^2\left[\gamma\left(C_y^2-2\rho C_y C_x+C_x^2\right)+o\left(n^{-2}\right)\right] \qquad \square$$

บทพิสูจน์ทฤษฎีบทที่ 3

ตัวประมาณค่าแบบอัตราส่วนแยกเดิมในการเลือกตัวอย่างแบบแบ่งชั้นภูมิอย่างง่าย $\bar{y}_{sr}$ ใน (37) สามารถเขียนใหม่ได้เป็น $\bar{y}_{sr} = \sum_{h=1}^{k} w_h \bar{y}_{rh}$ เมื่อ $\bar{y}_{rh} = \frac{\bar{y}_h}{\bar{x}_h} \bar{X}_h$ ดังนั้น โดยใช้ความเอนเอียงและความแปรปรวนของตัวประมาณ $\bar{y}_r$ ในทฤษฎีบทที่ 2 จะสามารถเขียนความเอนเอียง ความแปรปรวน และค่าคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณ $\bar{y}_{sr}$ ได้โดยง่ายเท่ากับ

$$\begin{aligned}
\mathrm{Bias}(\bar{y}_{sr}) &\approx \sum_{h=1}^{k} w_h \gamma_h \bar{Y}_h \left(C_{xh}^2 - \rho_h C_{yh} C_{xh}\right) \\
\mathrm{V}(\bar{y}_{sr}) &\approx \sum_{h=1}^{k} w_h^2 \bar{Y}_h^2 \left[\gamma_h \left(C_{yh}^2 - 2\rho_h C_{yh} C_{xh} + C_{xh}^2\right) - \gamma_h^2 \left(C_{xh}^2 - \rho_h C_{yh} C_{xh}\right)^2 + o(n_h^{-2})\right] \\
\mathrm{MSE}(\bar{y}_{sr}) &\approx \sum_{h=1}^{k} w_h^2 \bar{Y}_h^2 \left[\gamma_h \left(C_{yh}^2 - 2\rho_h C_{yh} C_{xh} + C_{xh}^2\right) - \gamma_h^2 \left(C_{xh}^2 - \rho_h C_{yh} C_{xh}\right)^2 + o(n_h^2)\right] \\
&\quad + \left[\sum_{h=1}^{k} w_h \gamma_h \bar{Y}_h \left(C_{xh}^2 - \rho_h C_{yh} C_{xh}\right)\right]^2
\end{aligned}$$

เมื่อ $o(n_h^{-2}) = \frac{\theta_{1h}}{n_h^2}\left(4C_{12h} - 2C_{03h} - 2C_{21h}\right) + \frac{\theta_{2h}}{n_h^2}\left(3C_{02h}C_{20h} + 6C_{11h}^2 + 9C_{02h}^2 - 18C_{02h}C_{11h}\right)$ □

บทพิสูจน์ทฤษฎีบทที่ 4

ความเอนเอียงของตัวประมาณ $\bar{y}_{sr.s}$ สามารถเขียนได้จากคำนิยามของ $\bar{y}_{sr.s}$ ใน (41) เท่ากับ

$$\mathrm{Bias}(\bar{y}_{sr.s}) = \sum_{h=1}^{k} w_h \left[\frac{1+\gamma_h C_{xh}^2}{1+\gamma_h C_{yh}^2} \mathrm{E}(\bar{y}_{rh}) - \bar{Y}_h\right]$$

ซึ่งสามารถพิสูจน์ได้โดยง่ายเท่ากับ

$$\mathrm{Bias}(\bar{y}_{sr.s}) \approx \sum_{h=1}^{k} w_h \frac{1+\gamma_h C_{xh}^2}{1+\gamma_h C_{yh}^2} \mathrm{Bias}(\bar{y}_{rh}) - \sum_{h=1}^{k} \frac{w_h \gamma_h \bar{Y}_h \left(C_{yh}^2 - C_{xh}^2\right)}{1+\gamma_h C_{yh}^2}$$

เมื่อ $\mathrm{Bias}(\bar{y}_{rh}) \approx \gamma_h \bar{Y}_h \left(C_{xh}^2 - \rho_h C_{yh} C_{xh}\right)$

ความแปรปรวนของตัวประมาณ $\bar{y}_{sr.s}$ สามารถเขียนได้จากคำนิยามของ $\bar{y}_{sr.s}$ ใน (41) เช่นเดียวกันเท่ากับ

$$\mathrm{V}(\bar{y}_{sr.s}) = \sum_{h=1}^{k} w_h^2 \mathrm{V}\left(\frac{1+\gamma_h C_{xh}^2}{1+\gamma_h C_{yh}^2} \bar{y}_{rh}\right) \approx \sum_{h=1}^{k} w_h^2 \left(\frac{1+\gamma_h C_{xh}^2}{1+\gamma_h C_{yh}^2}\right)^2 \mathrm{V}(\bar{y}_{rh})$$

และค่าคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณ $\bar{y}_{st.rs}$ เท่ากับ

$$\mathrm{MSE}(\bar{y}_{sr.s}) \approx \sum_{h=1}^{k} w_h^2 \left(\frac{1+\gamma_h C_{xh}^2}{1+\gamma_h C_{yh}^2}\right)^2 \mathrm{V}(\bar{y}_{rh}) + \left[\sum_{h=1}^{k} w_h \frac{1+\gamma_h C_{xh}^2}{1+\gamma_h C_{yh}^2} \mathrm{Bias}(\bar{y}_{rh}) - \sum_{h=1}^{k} \frac{w_h \gamma_h \bar{Y}_h \left(C_{yh}^2 - C_{xh}^2\right)}{1+\gamma_h C_{yh}^2}\right]^2$$

ซึ่งเมื่อเก็บเฉพาะพจน์ของ $1/n_h$ ไม่เกินองศาที่สอง จะเหลือเพียง

$$\mathrm{MSE}(\bar{y}_{sr.s}) \approx \sum_{h=1}^{k} w_h^2 \left(\frac{1+\gamma_h C_{xh}^2}{1+\gamma_h C_{yh}^2}\right)^2 \mathrm{V}(\bar{y}_{rh}) + \left[\sum_{h=1}^{k} w_h \frac{1+\gamma_h C_{xh}^2}{1+\gamma_h C_{yh}^2} \mathrm{Bias}(\bar{y}_{rh})\right]^2$$

$$-\sum_{h=1}^{k}\frac{w_h\gamma_h\bar{Y}_h\left(C_{yh}^2-C_{xh}^2\right)}{1+\gamma_hC_{yh}^2}\left[2\sum_{h=1}^{k}w_h\frac{1+\gamma_hC_{xh}^2}{1+\gamma_hC_{yh}^2}\mathrm{Bias}\left(\bar{y}_{rh}\right)-\sum_{h=1}^{k}\frac{w_h\gamma_h\bar{Y}_h\left(C_{yh}^2-C_{xh}^2\right)}{1+\gamma_hC_{yh}^2}\right]$$

เนื่องจาก $\mathrm{MSE}\left(\bar{y}_{sr.s}\right)$ มีเฉพาะพจน์ของ $1/n_h$ ไม่เกินองศาที่สอง ทำให้ $\mathrm{Bias}\left(\bar{y}_{rh}\right)$ ในผลบวกสุดท้าย จึงแทนด้วยการประมาณถึงอันดับที่หนึ่ง ก็เพียงพอ $\mathrm{Bias}\left(\bar{y}_{rh}\right)\approx\gamma_h\bar{Y}_h\left(C_{xh}^2-\rho_hC_{yh}C_{xh}\right)$ ดังนั้นจะได้ $\mathrm{MSE}\left(\bar{y}_{sr.s}\right)$ เท่ากับ

$$\mathrm{MSE}\left(\bar{y}_{sr.s}\right)\approx\sum_{h=1}^{k}w_h^2\left(\frac{1+\gamma_hC_{xh}^2}{1+\gamma_hC_{yh}^2}\right)^2\mathrm{V}\left(\bar{y}_{rh}\right)+\left[\sum_{h=1}^{k}w_h\frac{1+\gamma_hC_{xh}^2}{1+\gamma_hC_{yh}^2}\mathrm{Bias}\left(\bar{y}_{rh}\right)\right]^2$$
$$-\sum_{h=1}^{k}\frac{w_h\gamma_h\bar{Y}_h\left(C_{yh}^2-C_{xh}^2\right)}{1+\gamma_hC_{yh}^2}\left\{\sum_{h=1}^{k}\frac{w_h\gamma_h\bar{Y}_h}{1+\gamma_hC_{yh}^2}\left[2\gamma_hC_{xh}^4+3C_{xh}^2-C_{yh}^2-2\rho_hC_{yh}C_{xh}\left(1+\gamma_hC_{xh}^2\right)\right]\right\}$$

เมื่อ $\mathrm{V}\left(\bar{y}_{rh}\right)\approx\bar{Y}_h^2\left[\gamma_h\left(C_{yh}^2-2\rho_hC_{yh}C_{xh}+C_{xh}^2\right)-\gamma_h^2\left(C_{xh}^2-\rho_hC_{yh}C_{xh}\right)^2+o\left(n_h^{-2}\right)\right]$ □

บทพิสูจน์บทตั้งที่ 2

ตัวประมาณ $\bar{y}_{sr.s}$ จะมีประสิทธิภาพสูงกว่าตัวประมาณ $\bar{y}_{sr}$ เมื่อ $\mathrm{MSE}\left(\bar{y}_{sr}\right)>\mathrm{MSE}\left(\bar{y}_{sr.s}\right)$

$$\sum_{h=1}^{k}w_h^2\mathrm{V}\left(\bar{y}_{rh}\right)+\left[\sum_{h=1}^{k}w_h\mathrm{Bias}\left(\bar{y}_{rh}\right)\right]^2>\sum_{h=1}^{k}w_h^2\left(\frac{1+\gamma_hC_{xh}^2}{1+\gamma_hC_{yh}^2}\right)^2\mathrm{V}\left(\bar{y}_{rh}\right)+\left[\sum_{h=1}^{k}w_h\frac{1+\gamma_hC_{xh}^2}{1+\gamma_hC_{yh}^2}\mathrm{Bias}\left(\bar{y}_{rh}\right)\right]^2$$
$$-\sum_{h=1}^{k}\frac{w_h\gamma_h\bar{Y}_h\left(C_{yh}^2-C_{xh}^2\right)}{1+\gamma_hC_{yh}^2}\left\{\sum_{h=1}^{k}\frac{w_h\gamma_h\bar{Y}_h}{1+\gamma_hC_{yh}^2}\left[2\gamma_hC_{xh}^4+3C_{xh}^2-C_{yh}^2-2\rho_hC_{yh}C_{xh}\left(1+\gamma_hC_{xh}^2\right)\right]\right\}$$

พิจารณา $\mathrm{MSE}\left(\bar{y}_{sr.s}\right)$ โดยกระจายอนุกรมเทย์เลอร์ของ $\left(1+\gamma_hC_{yh}^2\right)^{-1}$ และ $\left(1+\gamma_hC_{yh}^2\right)^{-2}$ ถึงอันดับที่สอง จะได้

$$\mathrm{MSE}\left(\bar{y}_{sr.s}\right)\approx\mathrm{MSE}\left(\bar{y}_{sr}\right)-\left\{2\sum_{h=1}^{k}w_h^2\gamma_h^2\bar{Y}_h^2\left(C_{yh}^2-C_{xh}^2\right)\left(C_{yh}^2-2\rho_hC_{yh}C_{xh}+C_{xh}^2\right)\right.$$
$$\left.+\sum_{h=1}^{k}w_h\gamma_h\bar{Y}_h\left(C_{yh}^2-C_{xh}^2\right)\left[\sum_{h=1}^{k}w_h\gamma_h\bar{Y}_h\left(3C_{xh}^2-2\rho_hC_{yh}C_{xh}-C_{yh}^2\right)\right]\right\} \qquad (ก.7)$$

จาก (ก.7) จะเห็นว่า ตัวประมาณ $\bar{y}_{sr.s}$ จะมีประสิทธิภาพสูงกว่าตัวประมาณ $\bar{y}_{sr}$ ถ้า

$$2\sum_{h=1}^{k}w_h^2\gamma_h^2\bar{Y}_h^2\left(C_{yh}^2-C_{xh}^2\right)\left(C_{yh}^2-2\rho_hC_{yh}C_{xh}+C_{xh}^2\right)$$
$$+\left[\sum_{h=1}^{k}w_h\gamma_h\bar{Y}_h\left(C_{yh}^2-C_{xh}^2\right)\right]\left[\sum_{h=1}^{k}w_h\gamma_h\bar{Y}_h\left(3C_{xh}^2-2\rho_hC_{yh}C_{xh}-C_{yh}^2\right)\right]>0$$

$$\left[\sum_{h=1}^{k}w_h\gamma_h\bar{Y}_h\left(C_{yh}+C_{xh}\right)\left(C_{yh}-C_{xh}\right)\right]\left[2w_h\gamma_h\bar{Y}_h\left(C_{yh}^2-2\rho_hC_{yh}C_{xh}+C_{xh}^2\right)\right.$$
$$\left.+\sum_{h=1}^{k}w_h\gamma_h\bar{Y}_h\left(3C_{xh}^2-2\rho_hC_{yh}C_{xh}-C_{yh}^2\right)\right]>0$$

ถ้าทุกชั้นภูมิมีค่า $C_{yh}-C_{xh}>0$ เงื่อนไขที่ตัวประมาณ $\bar{y}_{sr.s}$ จะมีประสิทธิภาพสูงกว่าตัวประมาณ $\bar{y}_{sr}$ จะเหลือเพียง

$$\min_{1\le j\le k}w_j\gamma_j\bar{Y}_j\left(C_{yj}^2-2\rho_jC_{yj}C_{xj}+C_{xj}^2\right)>-\frac{1}{2}\sum_{h=1}^{k}w_h\gamma_h\bar{Y}_h\left(3C_{xh}^2-2\rho_hC_{yh}C_{xh}-C_{yh}^2\right)$$ □