# การประมาณค่าเฉลี่ยประชากรเมื่อมีข้อมูลสูญหาย ในการเลือกตัวอย่างสุ่มแบบแบ่งชั้นภูมิ

# Estimation of Population Mean with Missing Data in Stratified Random Sampling

กรกช ศิลปกอบ* และ วัชรีภรณ์ ไชยมงคล

Korakoch Silpakob* and Watchareephorn Chaimongkol

*คณะสถิติประยุกต์ สถาบันบัณฑิตพัฒนาบริหารศาสตร์*

*School of Applied Statistics, National Institute of Development Administration*

Received : 12 March 2017

Accepted : 8 June 2017

Published online : 22 June 2017

## บทคัดย่อ

งานวิจัยนี้ได้เสนอตัวประมาณค่าเฉลี่ยประชากรเมื่อมีข้อมูลสูญหายในการเลือกตัวอย่างสุ่มแบบแบ่งชั้นภูมิ ซึ่งพัฒนาจากตัวประมาณค่าเฉลี่ยประชากรเมื่อมีข้อมูลสูญหายในการเลือกตัวอย่างสุ่มแบบง่าย ของ Gira (2015) โดยการจำลองประชากรขนาด 150,000 หน่วย ที่มีตัวแปรที่สนใจ (Y) ตัวแปรช่วย (X) และกำหนดสัมประสิทธิ์สหสัมพันธ์ระหว่าง X กับ Y 3 ระดับ คือ 0.50, 0.75 และ 0.90 ซึ่งจะสุ่มตัวอย่างขนาด 30, 90, 300 และ 600 หน่วยตัวอย่าง และสุ่มเลือกให้ตัวแปรที่สนใจ (Y) มีเปอร์เซ็นต์การสูญหายอย่างสุ่มที่ 5%, 10% และ 15% ตามลำดับ เกณฑ์ที่ใช้ในการเปรียบเทียบ คือ ค่าคลาดเคลื่อนกำลังสองเฉลี่ย ผลการศึกษาพบว่า ตัวประมาณที่นำเสนอมีประสิทธิภาพมากกว่าตัวประมาณค่าเฉลี่ยประชากรเมื่อมีข้อมูลสูญหายด้วยค่าอัตราส่วนของ Singh *et al.* (2015) และตัวประมาณค่าเฉลี่ยประชากรเมื่อมีข้อมูลสูญหายของ Thakur *et al.* (2014) ในทุกกรณี

**คำสำคัญ** : ข้อมูลสูญหาย การเลือกตัวอย่างสุ่มแบบแบ่งชั้นภูมิ การเลือกตัวอย่างสุ่มแบบง่าย ตัวประมาณค่าเฉลี่ยประชากร ค่าคลาดเคลื่อนกำลังสองเฉลี่ย

## Abstract

This research presents estimator of population mean with missing data in stratified random sampling developed from the estimator of population mean with missing data in stratified random sampling concept of Gira (2015) by simulation. The size of the population is 150,000 units with values of a study variable Y and an auxiliary variable X. When the correlation coefficient between variable Y and X has 3 levels which is 0.05, 0.75 and 0.90. Random sample is 30, 90, 300 and 600 sample units are drawn from the population. From each sample randomly the variable Y as randomly missing at 5%, 10% and 15%, respectively. The criteria used for the comparison is the mean square error. In the present study, it indicates that the proposed estimator to be more effective than those reported in Singh *et al.* (2015) and Thakur *et al.* (2014) for every cases.

**Keywords** : missing data, stratified random sampling, simple random sampling, estimator of population mean, mean square error

***Corresponding author. E-mail** : korakoch14737@gmail.com

**บทนำ**

การเลือกตัวอย่างแบบแบ่งชั้นภูมิ (Stratified Sampling) เป็นวิธีการเลือกตัวอย่างที่ใช้กันแพร่หลายมากที่สุดไม่ว่าจะเป็นการสำรวจขนาดใหญ่ ขนาดเล็ก งานวิจัยเชิงสำรวจที่ครอบคลุมหลายพื้นที่ บุคคลหลายกลุ่ม หลายอาชีพ มักใช้วิธีการเลือกตัวอย่างนี้ การเลือกตัวอย่างแบบแบ่งชั้นภูมิ ได้แก่ การเลือกตัวอย่างจากประชากรหนึ่ง โดยแบ่งประชากรออกเป็นส่วน ๆ แต่ละส่วนเรียกว่าชั้นภูมิ (Stratum) แล้วสุ่มหน่วยตัวอย่างจากแต่ละชั้นภูมิ ด้วยวิธีการแบบใดแบบหนึ่ง โดยอิสระจากการสุ่มจากชั้นภูมิอื่น ๆ การสุ่มหน่วยตัวอย่างจากชั้นภูมิต่าง ๆ นั้นอาจจะใช้แบบเดียวกันหรือแตกต่างกันก็ได้ เช่น เมื่อสุ่มตัวอย่างจากแต่ละชั้นภูมิ ด้วยวิธีการเลือกตัวอย่างสุ่มแบบง่ายจะเรียกวิธีการนั้นว่า การเลือกตัวอย่างสุ่มแบบแบ่งชั้นภูมิ (Stratified Random Sampling) เมื่อสุ่มตัวอย่างจากแต่ละชั้นภูมิ ด้วยวิธีการเลือกตัวอย่างแบบมีระบบ จะเรียกวิธีการนั้นว่า การเลือกตัวอย่างแบบแบ่งชั้นภูมิอย่างมีระบบ (Stratified Systematic Sampling) นอกจากวิธีการดังกล่าวแล้ว อาจมีงานสำรวจด้วยตัวอย่างบางโครงการที่ใช้การสุ่มตัวอย่างจากแต่ละชั้นภูมิวิธีอื่น ๆ อีกก็ได้ แต่โครงสร้างของตัวอย่างจะคงใช้หลักการเดิมคือมาจากตัวอย่างที่ได้จากแต่ละชั้นภูมิโดยอิสระกัน การใช้วิธีการเลือกตัวอย่างแบบแบ่งชั้นภูมิจะให้ค่าประมาณของพารามิเตอร์ที่ได้มีความถูกต้องหรือความแม่นยำสูงกว่าวิธีอื่น ๆ ที่ใช้ตัวอย่างขนาดเท่า ๆ กัน หรือเสียค่าใช้จ่ายเท่า ๆ กัน สามารถทำการวิเคราะห์ข้อมูลเฉพาะบางส่วนของประชากร คือ วิเคราะห์แยกเป็นชั้นภูมิได้ ทำให้ได้สารสนเทศในแต่ละชั้นภูมิตามที่ต้องการ สามารถให้น้ำหนักความสำคัญแก่หน่วยตัวอย่างบางหน่วยได้สูงกว่าหน่วยอื่น ๆ เช่น ครัวเรือนที่มีรายได้สูงที่มีอยู่น้อยกว่าครัวเรือนที่มีรายได้ต่ำ แต่อาจมีความสำคัญสูงกว่าเมื่อต้องการศึกษารายได้ของประชากร โดยการให้น้ำหนักรวมสูงกว่าปกติ และสามารถใช้สารสนเทศบางประการที่มีอยู่ในบางชั้นภูมิให้เกิดประโยชน์ในการสุ่มตัวอย่างและในการประมาณค่าพารามิเตอร์ ซึ่งอาจทำให้ค่าประมาณแม่นยำมากยิ่งขึ้น (suwattee, 2009)

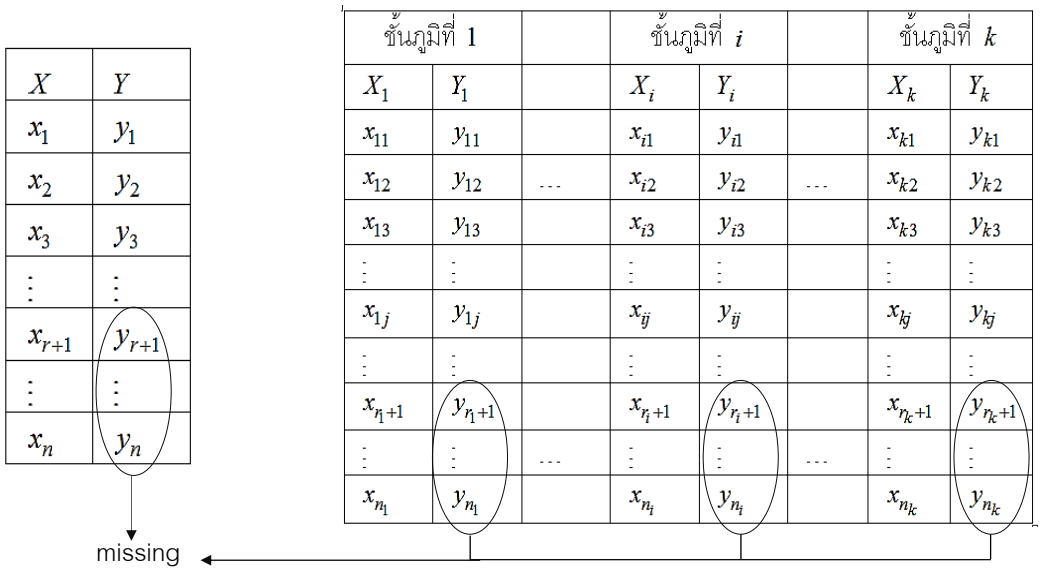
ข้อมูลสูญหาย (Missing Data) เป็นปัญหาที่พบบ่อยมากในการเก็บรวบรวมข้อมูลจากแบบสอบถามในการสำรวจด้วยตัวอย่าง ข้อมูลสูญหายจากการสำรวจด้วยตัวอย่างโดยทั่วไปมี 2 ประเภท คือ การไม่ตอบของหน่วยตัวอย่างบางหน่วย (Unit Nonresponse) และการไม่ตอบบางคำถามหรือบางตัวแปร (Item Nonresponse) ซึ่ง Kalton and Kasprzyk (1982) ได้ให้นิยามของการไม่ตอบของหน่วยตัวอย่างบางหน่วย คือการไม่ตอบของหน่วยตัวอย่างบางหน่วยซึ่งอาจเป็นผลสืบเนื่องมาจากการไม่เข้าใจความหมายของคำต่าง ๆ ที่ใช้ในแบบสอบถาม และการไม่ตอบบางคำถามหรือบางตัวแปร คือการสูญหายของข้อมูลที่เกิดจากการไม่ตอบเฉพาะบางคำถามหรือบางตัวแปร ซึ่งอาจเกิดจากการออกแบบคำถามในแบบสอบถามไม่ครอบคลุมทำให้ผู้ตอบแบบสอบถามไม่สามารถตอบคำถามบางคำถามได้ หรืออาจเกิดจากความผิดพลาดในการบันทึกข้อมูลทำให้ต้องตัดข้อมูลออกไป การจัดการกับข้อมูลสูญหายมีหลายวิธี ขึ้นอยู่กับลักษณะของข้อมูลสูญหายที่เกิดขึ้น ส่วนใหญ่การแก้ไขปัญหาข้อมูลสูญหายที่เกิดจากการไม่ตอบเฉพาะบางคำถามหรือบางตัวแปร (Item Nonresponse) ในทางสถิติจะมักใช้วิธีการประมาณค่าข้อมูลสูญหายด้วยค่าประมาณจากการคำนวณ (Imputation) แล้วนำค่าประมาณที่ได้ไปแทนค่าข้อมูลสูญหายเพื่อให้ได้ชุดข้อมูลที่สมบูรณ์เพื่อใช้ในการวิเคราะห์ต่อไป ซึ่งโดยทั่วไปจะแบ่งวิธีการประมาณข้อมูลสูญหายออกเป็น 2 กลุ่ม (Laaksonen, 2000) คือ 1) Model-donor Imputation คือการประมาณค่าที่ได้มาจากตัวแบบ 2) Real-donor Imputation คือการประมาณค่าที่ได้จากเซตข้อมูลของค่าที่สังเกตได้ ซึ่งงานวิจัยนี้สนใจศึกษาข้อมูลสูญหายที่เกิดจากการไม่ตอบบางคำถามหรือบางตัวแปร

การพัฒนาตัวประมาณค่าเฉลี่ยประชากรเมื่อมีข้อมูลสูญหายในงานวิจัยนี้พัฒนาจากตัวประมาณค่าเฉลี่ยประชากรเมื่อมีข้อมูลสูญหายในการเลือกตัวอย่างสุ่มแบบง่ายของ Gira (2015) โดยพัฒนาภายใต้วิธีการเลือกตัวอย่างสุ่มแบบแบ่งชั้นภูมิ และเปรียบเทียบประสิทธิภาพของตัวประมาณค่าเฉลี่ยประชากรเมื่อมีข้อมูลสูญหายในการเลือกตัวอย่างสุ่มแบบแบ่งชั้นภูมิที่พัฒนาขึ้นโดยใช้การจำลองข้อมูล และใช้ค่าคลาดเคลื่อนกำลังสองเฉลี่ยเป็นเกณฑ์ในการเปรียบเทียบประสิทธิภาพของตัวประมาณ

**ทบทวนทฤษฎีที่เกี่ยวข้อง**

ข้อมูลสูญหายคือค่าสังเกตที่ต้องการทราบค่าแต่ไม่สามารถทราบค่าได้หรือไม่สามารถเก็บรวบรวมได้โดยทั่วไปมี 2 ประเภท คือ ข้อมูลสูญหายที่เกิดจากการไม่ตอบของหน่วยตัวอย่างบางหน่วย (Unit Nonresponse) และข้อมูลสูญหายที่เกิดจากการไม่ตอบบางคำถามหรือบางตัวแปร (Item Nonresponse) ข้อมูลสูญหายที่ใช้ในงานวิจัยนี้คือข้อมูลสูญหายที่เกิดจากการไม่ตอบบางคำถามหรือบางตัวแปร ในกรณีข้อมูลสูญหายเกิดขึ้นในตัวแปรที่สนใจ ( $Y$ ) ซึ่งข้อมูลสูญหายที่เกิดขึ้นในตัวแปรที่สนใจมีลักษณะแสดงดังภาพที่ 1

**การเลือกตัวอย่างสุ่มแบบง่าย** **การเลือกตัวอย่างสุ่มแบบแบ่งชั้นภูมิ**




***ภาพที่ 1*** รูปแบบข้อมูลสูญหายที่เกิดขึ้นในตัวแปรที่สนใจ

พิจารณากรณีสุ่มตัวอย่างด้วยวิธีแบ่งประชากรขนาด $N$ ออกเป็น $k$ ชั้นภูมิ โดยที่ชั้นภูมิที่ $i$ มีหน่วยตัวอย่างอยู่ $N_i$ หน่วย $\sum_{i=1}^{k} N_i = N$ แล้วสุ่มตัวอย่างขนาด $n_i$ จากชั้นภูมิที่ $i$ ด้วยวิธีการเลือกตัวอย่างสุ่มแบบง่ายไม่คืนที่ โดยอิสระกัน $\sum_{i=1}^{k} n_i = n$ ค่าเฉลี่ยประชากร ( $\bar{Y}$ ) สามารถประมาณได้ดังนี้

$$\bar{y}_{st} = \sum_{i=1}^{k} W_i \bar{y}_i \quad (1)$$

โดยที่ $\bar{y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij}$

$W_i = \frac{N_i}{N}$ คือ น้ำหนักของชั้นภูมิที่ $i$

$y_{ij}$ คือ ค่าสังเกตของตัวแปรที่สนใจ $Y$ จากหน่วยตัวอย่างที่ $j$ ของตัวอย่าง ซึ่งสุ่มจากชั้นภูมิที่ $i$ ด้วยวิธีการเลือกตัวอย่างสุ่มแบบง่ายไม่คืนที่ เมื่อ $i = 1, \ldots, k$ และ $j = 1, 2, 3, \ldots, n_i$ (Suwattee, 2009)

**ตัวประมาณค่าเฉลี่ยประชากรเมื่อมีข้อมูลสูญหายในการเลือกตัวอย่างสุ่มแบบง่าย**

มีสัญกรณ์ที่ใช้ดังต่อไปนี้

$y_{.j}$ คือ ค่าประมาณข้อมูลสูญหาย เมื่อ $j = 1, \ldots, n$

$y_j$ คือ ค่าสังเกตของตัวแปรที่สนใจ $y$ เมื่อ $j = 1, \ldots, r$

$x_j$ คือ ค่าสังเกตของตัวแปรช่วย $x$ เมื่อ $j = 1, \ldots, n$

$\bar{x}_r$ คือ ค่าเฉลี่ยของตัวแปรช่วย $x$ ที่มีค่าของตัวแปรที่สนใจ $y$ สมบูรณ์ทั้งหมด $r$ หน่วยตัวอย่าง

$\bar{x}_n$ คือ ค่าเฉลี่ยของตัวแปรช่วย $x$ จากข้อมูลที่สมบูรณ์ทั้งหมด $n$ หน่วยตัวอย่าง

$\bar{y}_r$ คือ ค่าเฉลี่ยของตัวแปรที่สนใจ $y$ จากข้อมูลที่สมบูรณ์ทั้งหมด $r$ หน่วยตัวอย่าง

$C_Y$ คือ สัมประสิทธิ์การแปรผันของข้อมูลของตัวแปรที่สนใจ $Y$

$C_X$ คือ สัมประสิทธิ์การแปรผันของข้อมูลของตัวแปรช่วย $X$

$\rho$ คือ สัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรที่สนใจ $Y$ และตัวแปรช่วย $X$

$\bar{Y}$ คือ ค่าเฉลี่ยประชากรของตัวแปรที่สนใจ $Y$

$\bar{X}$ คือ ค่าเฉลี่ยประชากรของตัวแปรช่วย $X$

$r$ คือ จำนวนข้อมูลที่สมบูรณ์ทั้งหมด $r$ หน่วยตัวอย่าง ของหน่วยตัวอย่าง $n$ หน่วยตัวอย่าง

$n$ คือ จำนวนข้อมูลทั้งหมดของหน่วยตัวอย่าง $n$ หน่วยตัวอย่าง

$$C_X^2 = \frac{S_X^2}{\bar{X}^2}, \quad C_Y^2 = \frac{S_Y^2}{\bar{Y}^2}, \quad \rho = \frac{S_{XY}}{S_X S_Y}, \quad R = \frac{\bar{Y}}{\bar{X}}, \quad B = \frac{S_{XY}}{S_X^2}, \quad S_Y^2 = \frac{1}{N-1} \sum_{i=1}^{N} \left(Y_i - \bar{Y}\right)^2,$$

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

Gira (2015) ได้ศึกษาเรื่อง การประมาณค่าเฉลี่ยประชากรด้วยวิธีการประมาณค่าข้อมูลสูญหายแบบใหม่ โดยใช้ตัวประมาณแบบอัตราส่วน ซึ่งมีสมการประมาณค่าข้อมูลสูญหายดังนี้

$$y_{.j} = \begin{cases} y_j \\ \bar{y}_r \left[ n \left( \dfrac{\alpha - \bar{x}_r}{\alpha - \bar{x}_n} \right) - r \right] \dfrac{x_j}{\sum_{j=r+1}^{n} x_j} \end{cases} \quad (2)$$

เมื่อ $\alpha$ คือ ค่าคงตัวที่ทำให้ค่าคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณมีค่าต่ำสุด โดยที่ค่า $\alpha$ ที่เหมาะที่สุดหาได้จาก $\alpha = \dfrac{\bar{X}(R-B)}{B}$ เมื่อ $R = \dfrac{\bar{Y}}{\bar{X}}$ และ $B = \dfrac{S_{XY}}{S_X^2}$ ตัวประมาณค่าเฉลี่ยประชากรอยู่ในรูป

$$\bar{y}_G = \bar{y}_r \frac{\alpha - \bar{x}_r}{\alpha - \bar{x}_n} \quad (3)$$

ความเอนเอียงของตัวประมาณ $\bar{y}_G$ เท่ากับ

$$Bias(\bar{y}_G) = -\theta \bar{Y} \left( \frac{1}{r} - \frac{1}{n} \right) \left( \rho C_y C_x \right) \quad (4)$$

เมื่อ $\theta = \dfrac{\bar{X}}{\alpha - \bar{X}}$

ค่าคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณ $\bar{y}_G$ เท่ากับ

$$MSE(\bar{y}_G) \cong \left( \frac{1}{r} - \frac{1}{N} \right) S_Y^2 + \left( \frac{1}{r} - \frac{1}{n} \right) \left( \theta^2 R^2 S_X^2 - 2\theta R S_{XY} \right) \quad (5)$$

ค่าคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุดเท่ากับ

$$MSE(\bar{y}_G)_{\min} \cong \left( \frac{1}{r} - \frac{1}{n} \right) S_Y^2 - \left( \frac{1}{r} - \frac{1}{n} \right) \left( B^2 S_X^2 \right) \quad (6)$$

**ตัวประมาณค่าเฉลี่ยประชากรเมื่อมีข้อมูลสูญหายในการเลือกตัวอย่างสุ่มแบบแบ่งชั้นภูมิ**

มีสัญกรณ์ที่ใช้ดังต่อไปนี้

$y_{.ij}$ คือ ค่าประมาณข้อมูลสูญหายในชั้นภูมิที่ $i$ หน่วยตัวอย่างที่ $j$ เมื่อ $i = 1, \ldots, k$ และ $j = 1, \ldots, n_i$

$y_{ij}$ คือ ค่าสังเกตของตัวแปรที่สนใจ $y$ ในชั้นภูมิที่ $i$ หน่วยตัวอย่างที่ $j$

$x_{ij}$ คือ ค่าสังเกตของตัวแปรช่วย $x$ ในชั้นภูมิที่ $i$ หน่วยตัวอย่างที่ $j$

$\bar{y}_{r_i}$ คือ ค่าเฉลี่ยของตัวแปรที่สนใจ $y_{ij}$ จากข้อมูลที่สมบูรณ์ทั้งหมด $r_i$ หน่วยตัวอย่าง

$\bar{x}_{r_i}$ คือ ค่าเฉลี่ยของตัวแปรช่วย $x_i$ ที่มีค่าของตัวแปรที่สนใจ $y_i$ สมบูรณ์ทั้งหมด $r_i$ หน่วยตัวอย่าง

$\bar{x}_{n_i}$ คือ ค่าเฉลี่ยของตัวแปรช่วย $x_i$ จากข้อมูลที่สมบูรณ์ทั้งหมด $n_i$ หน่วยตัวอย่าง

$C_Y$ คือ สัมประสิทธิ์การแปรผันของข้อมูลของตัวแปรที่สนใจ $Y$

$C_X$ คือ สัมประสิทธิ์การแปรผันของข้อมูลของตัวแปรช่วย $X$

$\rho$ คือ สัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรที่สนใจ $Y$ และตัวแปรช่วย $X$

$\bar{Y}_{N_i}$ คือ ค่าเฉลี่ยประชากรของตัวแปรที่สนใจ $Y$ ในชั้นภูมิที่ $i$ เมื่อ $i = 1,\ldots,k$

$\bar{X}_{N_i}$ คือ ค่าเฉลี่ยประชากรของตัวแปรช่วย $X$ ในชั้นภูมิที่ $i$ เมื่อ $i = 1,\ldots,k$

$N_i$ คือ ขนาดของประชากรในชั้นภูมิที่ $i$ เมื่อ $i = 1,\ldots,k$

$r_i$ คือ จำนวนข้อมูลที่สมบูรณ์ทั้งหมด $r_i$ หน่วยตัวอย่าง ของตัวอย่าง $n_i$ หน่วยตัวอย่าง

$n_i$ คือ จำนวนข้อมูลทั้งหมดของตัวอย่าง $n_i$ หน่วยตัวอย่าง

$k$ คือ จำนวนชั้นภูมิ

$$W_i = \frac{N_i}{N}, \qquad L_i = E\left(\frac{1}{r_i}\right) = \left[\frac{1}{n_i f_{1i}} + \frac{(N_i - n_i)(1 - f_{1i})}{(N_i - 1) n_i^2 f_{1i}^2}\right], \qquad f_{1i} = \frac{r_i}{n_i},$$

$$R_i = \frac{\bar{Y}_{N_i}}{\bar{X}_{N_i}}, \qquad B_i = \frac{S_{X_i Y_i}}{S_{X_i}^2}, \qquad C_{X_i}^2 = \frac{S_{X_i}^2}{\bar{X}_{N_i}^2}, \qquad C_{Y_i}^2 = \frac{S_{Y_i}^2}{\bar{Y}_{N_i}^2}, \qquad \rho_i = \frac{S_{X_i Y_i}}{S_{X_i} S_{Y_i}},$$

$$S_{Y_i}^2 = \frac{1}{N_i - 1}\sum_{j=1}^{N_i}\left(Y_{ij} - \bar{Y}_{N_i}\right)^2, \qquad S_{X_i}^2 = \frac{1}{N_i - 1}\sum_{j=1}^{N_i}\left(X_{ij} - \bar{X}_{N_i}\right)^2,$$

$$S_{X_i Y_i} = \frac{1}{N_i - 1}\sum_{j=1}^{N_i}\left(X_{ij} - \bar{X}_{N_i}\right)\left(Y_{ij} - \bar{Y}_{N_i}\right)$$

Thakur *et al*. (2014) ได้ศึกษาเรื่องการประมาณค่าเฉลี่ยด้วยการประมาณค่าข้อมูลสูญหายในการเลือกตัวอย่างสุ่มแบบแบ่งชั้นภูมิ โดยได้เสนอและเปรียบเทียบตัวประมาณค่าเฉลี่ยประชากรโดยในการเลือกตัวอย่างสุ่มแบบแบ่งชั้นภูมิกับการเลือกตัวอย่างสุ่มแบบง่าย สำหรับตัวประประมาณค่าเฉลี่ยด้วยการประมาณค่าข้อมูลสูญหายในการเลือกตัวอย่างสุ่มแบบแบ่งชั้นภูมิ $\left(\bar{y}_{T_{st}}\right)$ มีสมการประมาณค่าข้อมูลสูญหายดังนี้

$$y_{.ij} = \begin{cases} y_{ij} & \text{ถ้า} \quad j = 1,\ldots,r_i \\ \dfrac{\bar{y}_{r_i}}{(1 - f_{1i})}\left[\left(\dfrac{\bar{X}_{n_i}}{\bar{x}_{r_i}}\right)^{\beta_{2i}} - f_{1i}\right] & \text{ถ้า} \quad j = r_i + 1,\ldots,n_i \end{cases} \tag{7}$$

เมื่อ $\beta_{2i} = \rho_i \dfrac{C_{Y_i}}{C_{X_i}}$

ตัวประมาณค่าเฉลี่ยประชากรอยู่ในรูป

$$\bar{y}_{T_{st}} = \sum_{i=1}^{k} W_i \bar{y}_{r_i}\left(\frac{\bar{X}_{n_i}}{\bar{x}_{r_i}}\right)^{\beta_{2i}} \tag{8}$$

ความเอนเอียงของตัวประมาณ $\bar{y}_{T_{st}}$ เท่ากับ

$$Bias\left(\bar{y}_{T_{st}}\right) = \sum_{i=1}^{k} W_i \bar{Y}_{N_i} \beta_{2i}\left(L_i - \frac{1}{n_i}\right)\left[\frac{(\beta_{2i} + 1)}{2} C_{X_i}^2 - p_i C_{Y_i} C_{X_i}\right] \tag{9}$$

ค่าคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณ $\bar{y}_{T_{st}}$ เท่ากับ

$$MSE\left(\bar{y}_{T_{st}}\right) = \sum_{i=1}^{k} W_i^{\,2} \bar{Y}_{N_i}^{2} \left[ \left( L_i - \frac{1}{N_i} \right) C_{Y_i}^{2} + \left( L_i - \frac{1}{n_i} \right) \left( \beta_{2i}^{2} C_{X_i}^{2} - 2\beta_{2i} \rho_i C_{Y_i} C_{X_i} \right) \right] \tag{10}$$

ค่าคลาดเคลื่อนกำลังสองเฉลี่ยที่เหมาะที่สุดเท่ากับ

$$MSE\left[\left(\bar{y}_{T_{st}}\right)\right]_{Min} = \sum_{i=1}^{k} W_i^{\,2} S_{Y_i}^{2} \left[ \left( L_i - \frac{1}{N_i} \right) - \left( L_i - \frac{1}{n_i} \right) \rho_i^{2} \right] \tag{11}$$

Singh *et al*. (2015) ได้ศึกษาเรื่อง การประมาณค่าเฉลี่ยประชากรภายใต้การไม่ตอบโดยใช้วิธีการประมาณค่าสูญหายต่าง ๆ สำหรับประชากรที่แบ่งชั้นภูมิ ซึ่งวิธีการประมาณค่าข้อมูลสูญหายแบบอัตราส่วนมีสมการประมาณค่าข้อมูลสูญหายดังนี้

$$y_{.ij} = \begin{cases} y_{ij} & \text{ถ้า} \quad j = 1, \ldots, r_i \\ \hat{b}_i x_{ij} & \text{ถ้า} \quad j = r_i + 1, \ldots, n_i \end{cases} \qquad \text{เมื่อ} \qquad \hat{b}_i = \frac{\bar{y}_{r_i}}{\bar{x}_{r_i}} \tag{12}$$

ตัวประมาณค่าเฉลี่ยประชากรอยู่ในรูป

$$\bar{y}_{R_{st}} = \sum_{i=1}^{k} W_i \bar{y}_{r_i} \frac{\bar{x}_{n_i}}{\bar{x}_{r_i}} \tag{13}$$

ความเอนเอียงของตัวประมาณ $\bar{y}_{R_{st}}$ เท่ากับ

$$Bias\left(\bar{y}_{R_{st}}\right) \cong \sum_{i=1}^{k} W_i \bar{Y}_{N_i} \left( L_i - \frac{1}{n_i} \right) \left( C_{X_i}^{2} - \rho_i C_{Y_i} C_{X_i} \right) \tag{14}$$

ค่าคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณ $\bar{y}_{R_{st}}$ เท่ากับ

$$MSE\left(\bar{y}_{R_{st}}\right) \cong \sum_{i=1}^{k} W_i^{\,2} \left[ \left( L_i - \frac{1}{N_i} \right) S_{Y_i}^{2} + \left( L_i - \frac{1}{n_i} \right) \left( R_i^{2} S_{X_i}^{2} - 2R_i S_{X_i Y_i} \right) \right] \tag{15}$$

**ผลการวิจัยและวิจารณ์ผล**

งานวิจัยนี้มีวัตถุประสงค์เพื่อพัฒนาตัวประมาณค่าเฉลี่ยประชากรเมื่อมีข้อมูลสูญหายของ Gira ในการเลือกตัวอย่างสุ่มแบบแบ่งชั้นภูมิ และเพื่อตรวจสอบตัวประมาณที่พัฒนาขึ้น จึงใช้ข้อมูลจากการจำลองโดยใช้เทคนิคมอนติคาร์โล (Monte Carlo Simulation) สร้างประชากรที่มีขนาด 150,000 หน่วย แบ่งเป็น 3 ชั้นภูมิ ที่ประกอบด้วยตัวแปรอิสระ ( $X$ ) และตัวแปรที่สนใจ ( $Y$ ) ที่มีการแจกแจงแบบปกติ และมีสัมประสิทธิ์สหสัมพันธ์ระหว่าง $X$ และ $Y$ 3 ระดับ คือ 0.50, 0.75 และ 0.90 และสุ่มตัวอย่างจากประชากรแต่ละชั้นภูมิด้วยวิธีการเลือกตัวอย่างสุ่มแบบง่ายไม่คืนที่ โดยกำหนดขนาดตัวอย่างเท่ากับ 30, 90, 300 และ 600 หน่วยตัวอย่าง หลังจากนั้นสุ่มตำแหน่งการสูญหายของข้อมูลให้กับตัวแปรที่สนใจ โดยกำหนดเปอร์เซ็นต์

การสูญหายเป็น 5% 10% และ 15% ในทุกขนาดตัวอย่าง และเปรียบเทียบประสิทธิภาพของตัวประมาณที่นำเสนอในรูปของค่าคลาดเคลื่อนกำลังสองเฉลี่ยพร้อมแสดงตัวอย่างการคำนวณเชิงตัวเลขเพื่อสนับสนุนผลที่ได้ในเชิงทฤษฎี ดังนี้

**ผลการวิจัยเชิงทฤษฎี**

ผู้วิจัยได้ศึกษาตัวประมาณค่าเฉลี่ยประชากรเมื่อมีข้อมูลสูญหายในการเลือกตัวอย่างสุ่มแบบง่าย ( $\bar{y}_G$ ) ของ Gira (2015) และได้เสนอตัวประมาณค่าเฉลี่ยประชากรเมื่อมีข้อมูลสูญหายในการเลือกตัวอย่างสุ่มแบบแบ่งชั้นภูมิ ( $\bar{y}_{G_{st}}$ ) โดยการปรับตัวประมาณ $\bar{y}_G$ ศึกษาภายใต้ข้อมูลสูญหายเกิดขึ้นในตัวแปรที่สนใจ ซึ่งผลการศึกษาเป็นดังนี้

**ตัวประมาณค่าเฉลี่ยประชากรเมื่อมีข้อมูลสูญหายในการเลือกตัวอย่างสุ่มแบบแบ่งชั้นภูมิที่นำเสนอ**

ตัวประมาณค่าเฉลี่ยประชากรเมื่อมีข้อมูลสูญหายในการเลือกตัวอย่างสุ่มแบบแบ่งชั้นภูมิ ( $\bar{y}_{G_{st}}$ ) มีสมการประมาณค่าสูญหายในการเลือกตัวอย่างสุ่มแบบแบ่งชั้นภูมิที่ประยุกต์จากแนวคิดของ Gira (2015) เป็นดังนี้

$$y_{.ij} = \begin{cases} y_{ij} & \text{ถ้า} \quad j = 1,\ldots,r_i \\ \bar{y}_{r_i}\left[ n_i\left( \dfrac{\alpha_i - \bar{x}_{r_i}}{\alpha_i - \bar{x}_{n_i}} \right) - r_i \right] \dfrac{x_{ij}}{\sum\limits_{j=r_i+1}^{n_i} x_{ij}} & \text{ถ้า} \quad j = r_i+1,\ldots,n_i \end{cases} \tag{16}$$

เมื่อ $\alpha_i$ คือ ค่าคงตัวที่ทำให้ค่าคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณค่ามีค่าต่ำสุด โดยที่ค่า $\alpha_i$ ที่เหมาะที่สุดหาได้จาก $\alpha_i = \dfrac{\bar{X}_i(R_i - B_i)}{B_i}$ เมื่อ $R_i = \dfrac{\bar{Y}_{N_i}}{\bar{X}_{N_i}}$ และ $B_i = \dfrac{S_{X_iY_i}}{S^2_{X_i}}$,

ตัวประมาณค่าเฉลี่ยประชากรเมื่อข้อมูลมีค่าสูญหายในการเลือกตัวอย่างสุ่มแบบแบ่งชั้นภูมิ( $\bar{y}_{G_{st}}$ ) ที่พัฒนามาจากตัวประมาณ $\bar{y}_G$ สามารถเขียนได้ดังนี้

$$\bar{y}_{G_{st}} = \sum_{i=1}^{k} W_i \bar{y}_{r_i} \frac{\alpha_i - \bar{x}_{r_i}}{\alpha_i - \bar{x}_{n_i}} \tag{17}$$

**คุณสมบัติของตัวประมาณ**

กำหนดให้ $\bar{y}_{r_i} = \bar{Y}_{N_i}(1+\varepsilon_i)$ , $\bar{x}_{r_i} = \bar{X}_{N_i}(1+\delta_i)$ และ $\bar{x}_{n_i} = \bar{X}_{N_i}(1+\eta_i)$

ใช้แนวคิด two-phase sampling (Rao and Sitter (1995) and Arnab and Singh (2006)) ภายใต้ข้อสมมติ ข้อมูลสูญหายแบบสุ่มอย่างสมบูรณ์ (MCAR) จะได้ว่า

$E(\varepsilon_i) = E(\delta_i) = E(\eta_i) = 0$ และ

$E(\varepsilon_i^2) = \left(\dfrac{1}{r_i} - \dfrac{1}{N_i}\right) C^2_{Y_i}$, $E(\delta_i^2) = \left(\dfrac{1}{r_i} - \dfrac{1}{N_i}\right) C^2_{X_i}$, $E(\varepsilon_i\delta_i) = \left(\dfrac{1}{r_i} - \dfrac{1}{N_i}\right) \rho_i C_{Y_i} C_{X_i}$,

$E(\eta_i^2) = \left(\dfrac{1}{n_i} - \dfrac{1}{N_i}\right) C^2_{X_i}$, $E(\delta_i\eta_i) = \left(\dfrac{1}{n_i} - \dfrac{1}{N_i}\right) C^2_{X_i}$, และ $E(\varepsilon_i\eta_i) = \left(\dfrac{1}{n_i} - \dfrac{1}{N_i}\right) \rho_i C_{Y_i} C_{X_i}$

**ทฤษฎีบทที่ 1** ตัวประมาณ $\bar{y}_{G_{st}}$ สามารถเขียนในเทอมของ $\varepsilon_i, \delta_i$ และ $\eta_i$ ได้ดังนี้

$$\bar{y}_{G_{st}} = \sum_{i=1}^{k} W_i \bar{Y}_{N_i} \left(1 + \varepsilon_i - \theta_i\delta_i + \theta_i\eta_i - \theta_i^2\eta_i\delta_i - \theta_i\varepsilon_i\delta_i + \theta_i\varepsilon_i\eta_i + \theta_i^2\eta_i^2 + O(\eta_i^2)\right) \tag{18}$$

**พิสูจน์**

$$\bar{y}_{G_{st}} = \sum_{i=1}^{k} W_i \bar{y}_{r_i} \frac{\alpha_i - \bar{x}_{r_i}}{\alpha_i - \bar{x}_{n_i}}$$

$$= \sum_{i=1}^{k} W_i \bar{Y}_{N_i} (1 + \varepsilon_i) \frac{\left(\frac{\bar{X}_{N_i}}{\theta_i} + \bar{X}_{N_i}\right) - \bar{X}_{N_i}(1 + \delta_i)}{\left(\frac{\bar{X}_{N_i}}{\theta_i} + \bar{X}_{N_i}\right) - \bar{X}_{N_i}(1 + \eta_i)}$$

$$= \sum_{i=1}^{k} W_i \bar{Y}_{N_i} (1 + \varepsilon_i) \frac{\left(\frac{1}{\theta_i} - \delta_i\right)}{\left(\frac{1}{\theta_i} - \eta_i\right)}$$

$$= \sum_{i=1}^{k} W_i \bar{Y}_{N_i} (1 + \varepsilon_i)(1 - \theta_i\delta_i)(1 - \theta_i\eta_i)^{-1}$$

$$= \sum_{i=1}^{k} W_i \bar{Y}_{N_i} (1 + \varepsilon_i)(1 - \theta_i\delta_i)\left(1 + \theta_i\eta_i + \theta_i^2\eta_i^2 + \ldots\right) \qquad (\text{จาก } |\theta_i\eta_i| < 1)$$

$$= \sum_{i=1}^{k} W_i \bar{Y}_{N_i} \left(1 + \theta_i\eta_i + \theta_i^2\eta_i^2 + \varepsilon_i + \theta_i\varepsilon_i\eta_i + \theta_i^2\varepsilon_i\eta_i^2 - \theta_i\delta_i - \theta_i^2\delta_i\eta_i - \theta_i^3\delta_i\eta_i^2 - \theta_i\varepsilon_i\delta_i - \theta_i^2\varepsilon_i\delta_i\eta_i - \theta^3\varepsilon_i\delta_i\eta_i^2 + \ldots\right)$$

$$= \sum_{i=1}^{k} W_i \bar{Y}_{N_i} \left(1 + \varepsilon_i - \theta_i\delta_i + \theta_i\eta_i - \theta_i^2\delta_i\eta_i - \theta_i\varepsilon_i\delta_i + \theta_i\varepsilon_i\eta_i + \theta_i^2\eta_i^2 + \left(\theta_i^2\varepsilon_i\eta_i^2 - \theta_i^3\delta_i\eta_i^2 - \theta_i^2\varepsilon_i\delta_i\eta_i - \theta^3\varepsilon_i\delta_i\eta_i^2 + \ldots\right)\right)$$

$$= \sum_{i=1}^{k} W_i \bar{Y}_{N_i} \left(1 + \varepsilon_i - \theta_i\delta_i + \theta_i\eta_i - \theta_i^2\delta_i\eta_i - \theta_i\varepsilon_i\delta_i + \theta_i\varepsilon_i\eta_i + \theta_i^2\eta_i^2 + O(\eta_i^2)\right)$$

**ทฤษฎีบทที่ 2** ความเอนเอียงของตัวประมาณ $\bar{y}_{G_{st}}$ เป็นดังนี้

$$Bias\left(\bar{y}_{G_{st}}\right) \cong -\sum_{i=1}^{k} \theta_i W_i \bar{Y}_{N_i} \left(\frac{1}{r_i} - \frac{1}{n_i}\right)\left(\rho_i C_{Y_i} C_{X_i}\right) \tag{19}$$

เมื่อ $\theta_i = \dfrac{\bar{X}_{N_i}}{\alpha_i - \bar{X}_{N_i}}$

**พิสูจน์** จากสมการที่ (18) เนื่องจากพจน์ของการกระจายเทย์เลอร์ (Taylor's expansion) อันดับที่มากกว่าหนึ่งเข้าใกล้ศูนย์ จึงสามารถประมาณความเอนเอียง ได้ดังนี้

$$Bias\left(\bar{y}_{G_{st}}\right) = E\left(\bar{y}_{G_{st}} - \bar{Y}_N\right)$$

$$\cong \sum_{i=1}^{k} W_i \bar{Y}_{N_i} E\left(1 + \varepsilon_i - \theta_i\delta_i + \theta_i\eta_i - \theta_i^2\delta_i\eta_i - \theta_i\varepsilon_i\delta_i + \theta_i\varepsilon_i\eta_i + \theta_i^2\eta_i^2 - \bar{Y}_N\right)$$

$$\cong \sum_{i=1}^{k} W_i \bar{Y}_{N_i} \left(-\theta_i^2 E(\delta_i\eta_i) - \theta_i E(\varepsilon_i\delta_i) + \theta_i E(\varepsilon_i\eta_i) + \theta_i^2 E(\eta_i^2)\right)$$

จากคุณสมบัติของตัวประมาณ $\bar{y}_{G_{st}}$ ตามแนวคิด two-phase sampling (Rao and Sitter (1995) and Arnab and Singh (2006)) ที่ $E(\delta_i\eta_i) = E(\eta_i^2)$ จะได้ $\theta_i^2 E(\eta_i^2) - \theta_i^2 E(\delta_i\eta_i) = 0$ จะได้ว่า

$$\cong \sum_{i=1}^{k} W_i \bar{Y}_{N_i} \left(-\theta_i E(\varepsilon_i\delta_i) + \theta_i E(\varepsilon_i\eta_i)\right)$$

$$\cong -\sum_{i=1}^{k} \theta_i W_i \bar{Y}_{N_i} \left[\left(\frac{1}{r_i} - \frac{1}{N_i}\right)\rho_i C_{Y_i} C_{X_i} - \left(\frac{1}{n_i} - \frac{1}{N_i}\right)\rho_i C_{Y_i} C_{X_i}\right]$$

$$\cong -\sum_{i=1}^{k} \theta_i W_i \bar{Y}_{N_i} \left(\frac{1}{r_i} - \frac{1}{n_i}\right)\left(\rho_i C_{Y_i} C_{X_i}\right)$$

**ทฤษฎีบทที่ 3** ค่าคลาดเคลื่อนกำลังสองเฉลี่ยที่เหมาะสมของตัวประมาณ $\bar{y}_{G_{st}}$ เป็นดังนี้

$$MSE(\bar{y}_{G_{st}})_{\min} \cong \sum_{i=1}^{k} W_i^2 \left[\left(\frac{1}{r_i} - \frac{1}{N_i}\right)S_{Y_i}^2 - \left(\frac{1}{r_i} - \frac{1}{n_i}\right)\left(B_i^2 S_{X_i}^2\right)\right] \tag{20}$$

**พิสูจน์** จากสมการที่ (18) จะได้ว่า

$$\bar{y}_{G_{st}} = \sum_{i=1}^{k} W_i \bar{Y}_{N_i} \left(1 + \varepsilon_i - \theta_i\delta_i + \theta_i\eta_i - \theta_i^2\eta_i\delta_i - \theta_i\varepsilon_i\delta_i + \theta_i\varepsilon_i\eta_i + \theta_i^2\eta_i^2 + O(\eta_i^2)\right)$$

$$\cong \sum_{i=1}^{k} W_i \bar{Y}_{N_i} \left(1 + \varepsilon_i - \theta_i\delta_i + \theta_i\eta_i\right)$$

$$\cong \sum_{i=1}^{k} W_i \bar{Y}_{N_i} + \sum_{i=1}^{k} W_i \bar{Y}_{N_i} \left(\varepsilon_i + \theta_i\eta_i - \theta_i\delta_i\right)$$

$$\bar{y}_{G_{st}} \cong \bar{Y}_N + \sum_{i=1}^{k} W_i \bar{Y}_{N_i} \left(\varepsilon_i + \theta_i(\eta_i - \delta_i)\right)$$

ดังนั้น

$$MSE\left(\bar{y}_{G_{st}}\right) = E\left(\bar{y}_{G_{st}} - \bar{Y}_N\right)^2$$

$$\cong E\left(\sum_{i=1}^{k} W_i \bar{Y}_{N_i} \left(\varepsilon_i + \theta_i(\eta_i - \delta_i)\right)\right)^2$$

$$\cong \sum_{i=1}^{k} W_i^2 \bar{Y}_{N_i}^2 E\left(\varepsilon_i + \theta_i(\eta_i - \delta_i)\right)^2$$

$$\cong \sum_{i=1}^{k} W_i^2 \bar{Y}_{N_i}^2 E\left(\varepsilon_i^2 + \theta_i^2(\eta_i - \delta_i)^2 + 2\theta_i(\varepsilon_i\eta_i - \varepsilon_i\delta_i)\right)$$

$$\cong \sum_{i=1}^{k} W_i^2 \bar{Y}_{N_i}^2 \left[E\left(\varepsilon_i^2\right) + \theta_i^2\left(E\left(\eta_i^2\right) + E\left(\delta_i^2\right) - 2E\left(\eta_i\delta_i\right)\right) + 2\theta_i\left(E\left(\varepsilon_i\eta_i\right) - E\left(\varepsilon_i\delta_i\right)\right)\right]$$

$$\cong \sum_{i=1}^{k} W_i^2 \bar{Y}_{N_i}^2 \left\{\left(\frac{1}{r_i} - \frac{1}{N_i}\right)C_{Y_i}^2 + \theta_i^2\left[\left(\frac{1}{n_i} - \frac{1}{N_i}\right)C_{X_i}^2 + \left(\frac{1}{r_i} - \frac{1}{N_i}\right)C_{X_i}^2 - 2\left(\frac{1}{n_i} - \frac{1}{N_i}\right)C_{X_i}^2\right] + 2\theta_i\left[\left(\frac{1}{n_i} - \frac{1}{N_i}\right)\rho_i C_{Y_i} C_{X_i} - \left(\frac{1}{r_i} - \frac{1}{N_i}\right)\rho_i C_{Y_i} C_{X_i}\right]\right\}$$

$$\cong \sum_{i=1}^{k} W_i^2 \bar{Y}_{N_i}^2 \left[\left(\frac{1}{r_i} - \frac{1}{N_i}\right)C_{Y_i}^2 + \left(\frac{1}{r_i} - \frac{1}{n_i}\right)\left(\theta_i^2 C_{X_i}^2 - 2\theta_i\rho_i C_{Y_i} C_{X_i}\right)\right]$$

$$MSE\left(\bar{y}_{G_{st}}\right) \cong \sum_{i=1}^{k} W_i^2 \left[\left(\frac{1}{r_i} - \frac{1}{N_i}\right)S_{Y_i}^2 + \left(\frac{1}{r_i} - \frac{1}{n_i}\right)\left(\theta_i^2 R_i^2 S_{X_i}^2 - 2\theta_i R_i S_{X_iY_i}\right)\right] \tag{21}$$

หาอนุพันธ์ลำดับที่ 1 เทียบกับ $\alpha_i$ ในสมการ (21) โดยการแทน $\theta_i = \frac{\bar{X}_{N_i}}{\alpha_i - \bar{X}_{N_i}}$ จะได้

$$MSE\left(\bar{y}_{G_{st}}\right) \cong \sum_{i=1}^{k} W_i^2 \left[\left(\frac{1}{r_i} - \frac{1}{N_i}\right) S_{Y_i}^2 + \left(\frac{1}{r_i} - \frac{1}{n_i}\right)\left(\theta_i^2 R_i^2 S_{X_i}^2 - 2\theta_i R_i S_{X_i Y_i}\right)\right]$$

$$\cong \sum_{i=1}^{k} W_i^2 \left[\left(\frac{1}{r_i} - \frac{1}{N_i}\right) S_{Y_i}^2 + \left(\frac{1}{r_i} - \frac{1}{n_i}\right)\left[\left(\frac{\bar{X}_{N_i}}{\alpha_i - \bar{X}_{N_i}}\right)^2 R_i^2 S_{X_i}^2 - 2\left(\frac{\bar{X}_{N_i}}{\alpha_i - \bar{X}_{N_i}}\right) R_i B_i S_{X_i}^2\right]\right]$$

$$\cong \sum_{i=1}^{k} W_i^2 \left\{\left(\frac{1}{r_i} - \frac{1}{N_i}\right) S_{Y_i}^2 + \left(\frac{1}{r_i} - \frac{1}{n_i}\right)\left[\frac{\left(\bar{X}_{N_i}\right)^2}{\left(\alpha_i - \bar{X}_{N_i}\right)^2} R_i^2 S_{X_i}^2 - 2\left(\frac{1}{r_i} - \frac{1}{n_i}\right)\left(\frac{\bar{X}_{N_i}}{\alpha_i - \bar{X}_{N_i}}\right) R_i B_i S_{X_i}^2\right]\right\}$$

$$\frac{\partial MSE\left(\bar{y}_{G_{st}}\right)}{\partial \alpha_i} \cong \frac{\partial}{\partial \alpha_i}\left\{\sum_{i=1}^{k} W_i^2 \left\{\left(\frac{1}{r_i} - \frac{1}{N_i}\right) S_{Y_i}^2 + \left(\frac{1}{r_i} - \frac{1}{n_i}\right)\left[\frac{\left(\bar{X}_{N_i}\right)^2}{\left(\alpha_i - \bar{X}_{N_i}\right)^2} R_i^2 S_{X_i}^2 - 2\left(\frac{1}{r_i} - \frac{1}{n_i}\right)\left(\frac{\bar{X}_{N_i}}{\alpha_i - \bar{X}_{N_i}}\right) R_i B_i S_{X_i}^2\right]\right\}\right\}$$

$$\cong \sum_{i=1}^{k} W_i^2 \left\{\left(\frac{1}{r_i} - \frac{1}{n_i}\right) R_i^2 S_{X_i}^2 \left[\frac{\left(\alpha_i - \bar{X}_{N_i}\right)^2 (0) - \bar{X}_{N_i}^2 2\left(\alpha_i - \bar{X}_{N_i}\right)}{\left(\alpha_i - \bar{X}_{N_i}\right)^4}\right]\right.$$

$$\left. - 2\left(\frac{1}{r_i} - \frac{1}{n_i}\right) R_i B_i S_{X_i}^2 \left[\frac{\left(\alpha_i - \bar{X}_{N_i}\right)(0) - \bar{X}_{N_i}(1)}{\left(\alpha_i - \bar{X}_{N_i}\right)^2}\right]\right\}$$

$$\cong \sum_{i=1}^{k} W_i^2 \left\{\left(\frac{1}{r_i} - \frac{1}{n_i}\right) R_i^2 S_{X_i}^2 \left[\frac{-2\bar{X}_{N_i}^2}{\left(\alpha_i - \bar{X}_{N_i}\right)^3}\right] - 2\left(\frac{1}{r_i} - \frac{1}{n_i}\right) R_i B_i S_{X_i}^2 \left[\frac{-\bar{X}_{N_i}}{\left(\alpha_i - \bar{X}_{N_i}\right)^2}\right]\right\}$$

และกำหนดให้เท่ากับ 0 จะได้

$$0 = \sum_{i=1}^{k} W_i^2 \left\{\left(\frac{1}{r_i} - \frac{1}{n_i}\right) R_i^2 S_{X_i}^2 \left[\frac{-2\bar{X}_{N_i}^2}{\left(\alpha_i - \bar{X}_{N_i}\right)^3}\right] - 2\left(\frac{1}{r_i} - \frac{1}{n_i}\right) R_i B_i S_{X_i}^2 \left[\frac{-\bar{X}_{N_i}}{\left(\alpha_i - \bar{X}_{N_i}\right)^2}\right]\right\}$$

$$0 = -2\left(\frac{1}{r_i} - \frac{1}{n_i}\right) R_i^2 S_{X_i}^2 \left[\frac{\bar{X}_{N_i}^2}{\left(\alpha_i - \bar{X}_{N_i}\right)^3}\right] + 2\left(\frac{1}{r_i} - \frac{1}{n_i}\right) R_i B_i S_{X_i}^2 \left[\frac{\bar{X}_{N_i}}{\left(\alpha_i - \bar{X}_{N_i}\right)^2}\right]$$

$$2\left(\frac{1}{r_i} - \frac{1}{n_i}\right) R_i^2 S_{X_i}^2 \left[\frac{\bar{X}_{N_i}^2}{\left(\alpha_i - \bar{X}_{N_i}\right)^3}\right] = 2\left(\frac{1}{r_i} - \frac{1}{n_i}\right) R_i B_i S_{X_i}^2 \left[\frac{\bar{X}_{N_i}}{\left(\alpha_i - \bar{X}_{N_i}\right)^2}\right]$$

$$R_i \left[\frac{\bar{X}_{N_i}}{\left(\alpha_i - \bar{X}_{N_i}\right)}\right] = B_i$$

$$\alpha_i = \frac{R_i \bar{X}_{N_i}}{B_i} + \bar{X}_{N_i}$$

$$\alpha_i = \frac{\bar{X}_{N_i}\left(R_i + B_i\right)}{B_i}$$

$$\alpha_i = \frac{\bar{X}_{N_i}(R_i + B_i)}{B_i}$$

แทน $\alpha_i = \frac{\bar{X}_{N_i}(R_i + B_i)}{B_i}$ ในสมการ $\theta_i = \frac{\bar{X}_{N_i}}{\alpha_i - \bar{X}_{N_i}}$ จะได้ $R_i\theta_i = B_i \Rightarrow \theta_i = \frac{B_i}{R_i}$ แทน $\theta_i = \frac{B_i}{R_i}$ ใน (21) จะได้

$$MSE\left(\bar{y}_{G_{st}}\right)_{\min} \cong \sum_{i=1}^{k} W_i^2 \left[\left(\frac{1}{r_i} - \frac{1}{N_i}\right) S_{Y_i}^2 + \left(\frac{1}{r_i} - \frac{1}{n_i}\right)\left(\left(\frac{B_i}{R_i}\right)^2 R_i^2 S_{X_i}^2 - 2\left(\frac{B_i}{R_i}\right) R_i S_{X_iY_i}\right)\right]$$

$$\cong \sum_{i=1}^{k} W_i^2 \left[\left(\frac{1}{r_i} - \frac{1}{N_i}\right) S_{Y_i}^2 + \left(\frac{1}{r_i} - \frac{1}{n_i}\right)\left(B_i^2 S_{X_i}^2 - 2B_i S_{X_iY_i}\right)\right]$$

$$\cong \sum_{i=1}^{k} W_i^2 \left[\left(\frac{1}{r_i} - \frac{1}{N_i}\right) S_{Y_i}^2 + \left(\frac{1}{r_i} - \frac{1}{n_i}\right)\left(B_i^2 S_{X_i}^2 - 2B_i\left(B_i S_{X_i}^2\right)\right)\right]$$

$$\cong \sum_{i=1}^{k} W_i^2 \left[\left(\frac{1}{r_i} - \frac{1}{N_i}\right) S_{Y_i}^2 + \left(\frac{1}{r_i} - \frac{1}{n_i}\right)\left(B_i^2 S_{X_i}^2 - 2B_i^2 S_{X_i}^2\right)\right]$$

$$\cong \sum_{i=1}^{k} W_i^2 \left[\left(\frac{1}{r_i} - \frac{1}{N_i}\right) S_{Y_i}^2 - \left(\frac{1}{r_i} - \frac{1}{n_i}\right) B_i^2 S_{X_i}^2\right]$$

$$MSE\left(\bar{y}_{G_{st}}\right)_{\min} \cong \sum_{i=1}^{k} W_i^2 \left[\left(\frac{1}{r_i} - \frac{1}{N_i}\right) S_{Y_i}^2 - \left(\frac{1}{r_i} - \frac{1}{n_i}\right) B_i^2 S_{X_i}^2\right]$$

**ผลการวิจัยเชิงตัวเลขและการวิจารณ์ผลการวิจัย**

ในหัวข้อนี้มีวัตถุประสงค์เพื่อแสดงตัวอย่างการคำนวณเชิงตัวเลขเพื่อเปรียบเทียบประสิทธิภาพของตัวประมาณค่าเฉลี่ยประชากรเมื่อมีข้อมูลสูญหายในการเลือกตัวอย่างสุ่มแบบแบ่งชั้นภูมิ ทั้ง 3 ตัวประมาณ ด้วยค่าคลาดคาดเคลื่อนกำลังสองเฉลี่ย ตัวประมาณค่าที่มีค่าคลาดเคลื่อนกำลังสองเฉลี่ยน้อยที่สุดแสดงว่า มีประสิทธิภาพมากที่สุด

โดยที่ $\bar{y}_{T_{st}}$ คือ ตัวประมาณค่าเฉลี่ยประชากรเมื่อมีข้อมูลสูญหายในการเลือกตัวอย่างสุ่มแบบแบ่งชั้นภูมิของ Thakur *et al*. (2014)

$\bar{y}_{R_{st}}$ คือ ตัวประมาณค่าเฉลี่ยประชากรเมื่อมีข้อมูลสูญหายในการเลือกตัวอย่างสุ่มแบบแบ่งชั้นภูมิของ Singh *et al*. (2015)

$\bar{y}_{G_{st}}$ คือ ตัวประมาณค่าเฉลี่ยประชากรเมื่อมีข้อมูลสูญหายในการเลือกตัวอย่างสุ่มแบบแบ่งชั้นภูมิที่นำเสนอ

$n$ คือ ขนาดตัวอย่าง

$m$ คือ ขนาดของข้อมูลสูญหายเป็นเปอร์เซ็นต์

**ข้อมูลชุดที่ 1** ประชากรใช้ข้อมูลจากการจำลองสถานการณ์ด้วยโปแกรม SAS จำลองประชากรขนาด 150,000 หน่วย ซึ่งประกอบด้วยตัวแปรที่สนใจ (Y) ตัวแปรช่วย (X) ที่สร้างขึ้น ณ ระดับค่าสัมประสิทธิ์สหสัมพันธ์ (correlation coefficient) เท่ากับ 0.50 ลักษณะของประชากรของตัวแปรช่วย $\bar{X}$ = 320 ตัวแปรที่สนใจ $\bar{Y}$ = 640

***ตารางที่ 1*** ค่าเอนเอียง ค่าคลาดเคลื่อนกำลังสองเฉลี่ยและค่าความแปรปรวนของตัวประมาณค่า $\bar{y}_{T_{st}}$, $\bar{y}_{R_{st}}$ และ $\bar{y}_{G_{st}}$ โดยมีค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ 0.50

| $n$ | $m$ | $\bar{y}_{T_{st}}$ | | | $\bar{y}_{R_{st}}$ | | | $\bar{y}_{G_{st}}$ | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | \|Bias\| | MSE | Variance | \|Bias\| | MSE | Variance | \|Bias\| | MSE | Variance |
| 30 | 5 | 0.0036 | 285.7330 | 285.7331 | 0.0159 | 287.7620 | 287.7623 | 0.0565 | 283.7585 | 283.7617 |
| | 10 | 0.0085 | 295.6215 | 295.6215 | 0.0440 | 300.3040 | 300.3059 | 0.0727 | 292.4111 | 292.4163 |
| | 15 | 0.0132 | 321.9525 | 321.9527 | 0.0645 | 329.2312 | 329.2354 | 0.1443 | 315.7382 | 315.7590 |
| 90 | 5 | 0.0012 | 93.3245 | 93.3245 | 0.0060 | 93.9914 | 93.9914 | 0.0138 | 93.1489 | 93.1491 |
| | 10 | 0.0023 | 97.8087 | 97.8087 | 0.0113 | 99.0696 | 99.0697 | 0.0236 | 97.4681 | 97.4687 |
| | 15 | 0.0041 | 104.1315 | 104.1315 | 0.0211 | 106.4113 | 106.4118 | 0.0367 | 103.5167 | 103.5181 |
| 300 | 5 | 0.0003 | 27.7486 | 27.7486 | 0.0017 | 27.9352 | 27.9352 | 0.0033 | 27.7346 | 27.7346 |
| | 10 | 0.0007 | 29.2207 | 29.2207 | 0.0036 | 29.6100 | 29.6101 | 0.0066 | 29.1898 | 29.1899 |
| | 15 | 0.0011 | 30.8650 | 30.8650 | 0.0054 | 31.4639 | 31.4639 | 0.0106 | 30.8137 | 30.8138 |
| 600 | 5 | 0.0002 | 13.8444 | 13.8444 | 0.0008 | 13.9359 | 13.9359 | 0.0016 | 13.8410 | 13.8410 |
| | 10 | 0.0003 | 14.5754 | 14.5754 | 0.0017 | 14.7649 | 14.7649 | 0.0034 | 14.5677 | 14.5677 |
| | 15 | 0.0005 | 15.3934 | 15.3934 | 0.0028 | 15.6968 | 15.6968 | 0.0053 | 15.3805 | 15.3806 |

จากตารางที่ 1 เมื่อพิจารณาจากค่า MSE พบว่า ตัวประมาณ $\bar{y}_{G_{st}}$ มีค่าคาดเคลื่อนกำลังสองเฉลี่ยน้อยกว่าค่าคาดคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณ $\bar{y}_{T_{st}}$ และ $\bar{y}_{R_{st}}$ ในทุกขนาดตัวอย่างและทุกเปอร์เซ็นต์การสูญหาย ดังนั้นตัวประมาณค่า $\bar{y}_{ST}$ มีประสิทธิภาพมากกว่าตัวประมาณ $\bar{y}_{T_{st}}$ และ $\bar{y}_{R_{st}}$

**ข้อมูลชุดที่ 2** ประชากรใช้ข้อมูลจากการจำลองสถานการณ์ด้วยโปแกรม SAS จำลองประชากรขนาด 150,000 หน่วย ซึ่งประกอบด้วยตัวแปรที่สนใจ (Y) ตัวแปรช่วย (X) ที่สร้างขึ้น ณ ระดับค่าสัมประสิทธิ์สหสัมพันธ์ เท่ากับ 0.75 ลักษณะของประชากรของตัวแปรช่วย $\bar{X}$ = 320 ตัวแปรที่สนใจ $\bar{Y}$ = 640

***ตารางที่ 2*** ค่าเอนเอียง ค่าคลาดเคลื่อนกำลังสองเฉลี่ยและค่าความแปรปรวนของตัวประมาณค่า $\bar{y}_{T_{st}}$, $\bar{y}_{R_{st}}$ และ $\bar{y}_{G_{st}}$ โดยมีค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ 0.75

| $n$ | $m$ | $\bar{y}_{T_{st}}$ | | | $\bar{y}_{R_{st}}$ | | | $\bar{y}_{G_{st}}$ | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | \|Bias\| | MSE | Variance | \|Bias\| | MSE | Variance | \|Bias\| | MSE | Variance |
| 30 | 5 | 0.0033 | 194.0586 | 194.0586 | 0.0101 | 194.8822 | 194.8823 | 0.0367 | 192.8897 | 192.8911 |
| | 10 | 0.0083 | 199.7451 | 199.7452 | 0.0279 | 201.6441 | 201.6449 | 0.1276 | 197.8656 | 197.8819 |
| | 15 | 0.0126 | 215.3024 | 215.3026 | 0.0409 | 218.2529 | 218.2546 | 0.1737 | 211.6428 | 211.6730 |
| 90 | 5 | 0.0012 | 63.5908 | 63.5908 | 0.0038 | 63.8612 | 63.8613 | 0.0178 | 63.4875 | 63.4879 |
| | 10 | 0.0022 | 66.2121 | 66.2121 | 0.0072 | 66.7255 | 66.7255 | 0.0344 | 66.0124 | 66.0135 |
| | 15 | 0.0040 | 69.8950 | 69.8950 | 0.0134 | 70.8238 | 70.8240 | 0.0661 | 69.5357 | 69.5400 |
| 300 | 5 | 0.0003 | 18.9087 | 18.9087 | 0.0011 | 18.9846 | 18.9846 | 0.0055 | 18.9005 | 18.9005 |
| | 10 | 0.0007 | 19.7688 | 19.7688 | 0.0023 | 19.9276 | 19.9276 | 0.0115 | 19.7508 | 19.7509 |
| | 15 | 0.0010 | 20.7322 | 20.7322 | 0.0035 | 20.9764 | 20.9764 | 0.0173 | 20.7022 | 20.7025 |
| 600 | 5 | 0.0002 | 9.4338 | 9.4338 | 0.0005 | 9.4711 | 9.4711 | 0.0027 | 9.4318 | 9.4318 |
| | 10 | 0.0003 | 9.8619 | 9.8619 | 0.0011 | 9.9391 | 9.9391 | 0.0055 | 9.8574 | 9.8574 |
| | 15 | 0.0005 | 10.3404 | 10.3404 | 0.0018 | 10.4640 | 10.4640 | 0.0089 | 10.3328 | 10.3329 |

จากตารางที่ 2 เมื่อพิจารณาจากค่า MSE พบว่า ตัวประมาณ $\bar{y}_{G_{st}}$ มีค่าคาดเคลื่อนกำลังสองเฉลี่ยน้อยกว่าค่าคาดคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณ $\bar{y}_{T_{st}}$ และ $\bar{y}_{R_{st}}$ ในทุกขนาดตัวอย่างและทุกเปอร์เซ็นต์การสูญหาย ดังนั้นตัวประมาณค่า $\bar{y}_{ST}$ มีประสิทธิภาพมากกว่าตัวประมาณ $\bar{y}_{T_{st}}$ และ $\bar{y}_{R_{st}}$

**ข้อมูลชุดที่ 3** ประชากรใช้ข้อมูลจากการจำลองสถานการณ์ด้วยโปแกรม SAS จำลองประชากรขนาด 150,000 หน่วย ซึ่งประกอบด้วยตัวแปรที่สนใจ (Y) ตัวแปรช่วย (X) ที่สร้างขึ้น ณ ระดับค่าสัมประสิทธิ์สหสัมพันธ์ เท่ากับ 0.90 ลักษณะของประชากรของตัวแปรช่วย $\bar{X}$ = 320 ตัวแปรที่สนใจ $\bar{Y}$ = 640

***ตารางที่ 3*** ค่าเอนเอียง ค่าคลาดเคลื่อนกำลังสองเฉลี่ยและค่าความแปรปรวนของตัวประมาณค่า $\bar{y}_{T_{st}}$, $\bar{y}_{R_{st}}$ และ $\bar{y}_{G_{st}}$ โดยมีค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ 0.90

| $n$ | $m$ | $\bar{y}_{T_{st}}$ | | | $\bar{y}_{R_{st}}$ | | | $\bar{y}_{G_{st}}$ | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | \|Bias\| | MSE | Variance | \|Bias\| | MSE | Variance | \|Bias\| | MSE | Variance |
| 30 | 5 | 0.0026 | 120.3257 | 120.3257 | 0.0067 | 120.6919 | 120.6919 | 0.0297 | 119.8181 | 119.8190 |
| | 10 | 0.0068 | 122.7953 | 122.7954 | 0.0186 | 123.6395 | 123.6398 | 0.0721 | 121.9791 | 121.9843 |
| | 15 | 0.0101 | 129.5517 | 129.5518 | 0.0273 | 130.8633 | 130.8641 | 0.1095 | 127.9623 | 127.9743 |
| 90 | 5 | 0.0009 | 39.7095 | 39.7095 | 0.0025 | 39.8297 | 39.8297 | 0.0110 | 39.6646 | 39.6647 |
| | 10 | 0.0018 | 40.8479 | 40.8479 | 0.0048 | 41.0761 | 41.0761 | 0.0205 | 40.7611 | 40.7615 |
| | 15 | 0.0032 | 42.4473 | 42.4473 | 0.0089 | 42.8602 | 42.8603 | 0.0373 | 42.2913 | 42.2926 |
| 300 | 5 | 0.0003 | 11.8184 | 11.8184 | 0.0007 | 11.8522 | 11.8522 | 0.0032 | 11.8149 | 11.8149 |
| | 10 | 0.0005 | 12.1920 | 12.1920 | 0.0015 | 12.2626 | 12.2626 | 0.0065 | 12.1842 | 12.1842 |
| | 15 | 0.0008 | 12.6104 | 12.6104 | 0.0023 | 12.7189 | 12.7189 | 0.0100 | 12.5973 | 12.5974 |
| 600 | 5 | 0.0001 | 5.8967 | 5.8967 | 0.0004 | 5.9133 | 5.9133 | 0.0015 | 5.8958 | 5.8958 |
| | 10 | 0.0003 | 6.0826 | 6.0826 | 0.0007 | 6.1169 | 6.1169 | 0.0032 | 6.0807 | 6.0807 |
| | 15 | 0.0004 | 6.2904 | 6.2904 | 0.0012 | 6.3454 | 6.3454 | 0.0051 | 6.2871 | 6.2872 |

จากตารางที่ 3 เมื่อพิจารณาจากค่า MSE พบว่า ตัวประมาณ $\bar{y}_{G_{st}}$ มีค่าคาดเคลื่อนกำลังสองเฉลี่ยน้อยกว่าค่าคาดคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณ $\bar{y}_{T_{st}}$ และ $\bar{y}_{R_{st}}$ ในทุกขนาดตัวอย่างและทุกเปอร์เซ็นต์การสูญหาย ดังนั้นตัวประมาณ $\bar{y}_{G_{st}}$ มีประสิทธิภาพมากกว่าตัวประมาณ $\bar{y}_{T_{st}}$ และ $\bar{y}_{R_{st}}$

จากตารางที่ 1, 2 และ 3 จะสังเกตเห็นได้ว่า ความเอนเอียงของตัวประมาณ $\bar{y}_{G_{st}}$ มีค่ามากกว่าความเอนเอียงของตัวประมาณ $\bar{y}_{T_{st}}$ และ $\bar{y}_{R_{st}}$ แต่ความแปรปรวนและค่าคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณ $\bar{y}_{G_{st}}$ มีค่าน้อยกว่าค่าคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณค่า $\bar{y}_{T_{st}}$ และ $\bar{y}_{R_{st}}$ ซึ่งการนำตัวประมาณค่า $\bar{y}_{G_{st}}$ มาใช้ก็เพื่อยอมให้มีความเอนเอียงมากขึ้น เพื่อให้ได้ค่าคลาดเคลื่อนกำลังสองเฉลี่ยน้อยลงทำให้ตัวประมาณค่ามีประสิทธิภาพมากขึ้น

**สรุปผลการวิจัย**

ตัวประมาณค่าเฉลี่ยประชากรเมื่อมีข้อมูลสูญหายในการเลือกตัวอย่างสุ่มแบบแบ่งชั้นภูมิ ได้พัฒนามาจากตัวประมาณค่าเฉลี่ยประชากรเมื่อมีข้อมูลสูญหายในการเลือกตัวอย่างสุ่มแบบง่าย จากแนวคิดของ Gira (2015) โดยใช้ตัวประมาณแบบอัตราส่วน จากผลการศึกษาเชิงตัวเลขพบว่า ตัวประมาณค่าเฉลี่ยประชากรเมื่อมีข้อมูลสูญหายในการเลือกตัวอย่างสุ่มแบบแบ่งชั้นภูมิ ( $\bar{y}_{G_{st}}$ ) ที่นำเสนอมีประสิทธิภาพมากกว่าตัวประมาณค่าเฉลี่ยประชากรเมื่อมีข้อมูลสูญหายในการเลือกตัวอย่างสุ่มแบบแบ่งชั้นภูมิ ( $\bar{y}_{T_{st}}$ ) ของ Thakur *et al*. (2014) และตัวประมาณค่าเฉลี่ยประชากรเมื่อมีข้อมูลสูญหายในการเลือกตัวอย่างสุ่มแบบแบ่งชั้นภูมิ ( $\bar{y}_{R_{st}}$ ) ของ Singh *et al*. (2015) ในทุกกรณี

**ข้อเสนอแนะ**

งานวิจัยนี้ได้พัฒนาตัวประมาณค่าเฉลี่ยประชากรเมื่อมีข้อมูลสูญหายของ Gira (2015) ในการเลือกตัวอย่างสุ่มแบบแบ่งชั้นภูมิดังนั้นการวิจัยครั้งต่อไปอาจใช้แนวคิดของGiraในการพัฒนาตัวประมาณค่าเฉลี่ยประชากรเมื่อมีข้อมูลสูญหายในการเลือกตัวอย่างแบบแบ่งชั้นอย่างมีระบบ (Stratified Systematic Sampling) หรือใช้ในการพัฒนาตัวประมาณค่าเฉลี่ยประชากรเมื่อมีข้อมูลสูญหายในการเลือกตัวอย่างแบบอื่น เป็นต้น

**เอกสารอ้างอิง**


Arnab, R. and Singh, S. (2006). A New Method for Estimating Variance from Data Imputed with Ratio Method of Imputation. *Statistics & probability letters*, *76* , 513-519.

Gira, Abdeltawab A. (2015). Estimation of Population Mean with a New Imputation Method. *Applied Mathematical Sciences, 9* , 1663-1672.

Kalton, G. and Kasprzyk, D. (1982). Imputation for Missing Survey Responses. *Proceeding Section of Survey Research Method. American Statistical Association, 22,* 22-33.

Laaksonen, S. (2000). Regression-Based Nearest neighbor Hot Decking. *Computational Statistics, 15* , 65-66.

Rao, J. N. K. and Sitter, R.R. (1995). Variance Estimation under Two-Phase Sampling with Application to Imputation for Missing Data. *Biometrika Trust*, *82* , 453-460.

Thakur, N. S., Yadav, K. and Pathak, S. (2014). Estimation of Mean with Imputation of Missing Data in Stratified Random Sampling. *STM Journals, 3*, 1-12.

Singh, P., Singh, A. K. and Singh, V.K. (2015). Estimation of Population Mean under Non-Response using Various Imputation Methods for Stratified Population. *Columbia International Publishing*, *4,* 112-122.

Suwattee, P. (2009). *Sample Surveys: Sampling Designs and Analysis*. Bangkok: WVO Officer of Printing Mill. (in Thai)