การเปรียบเทียบประสิทธิภาพของตัวประมาณค่าต่าง ๆ กับตัวประมาณค่าของซิลส์ที่ปรับปรุงใหม่

วิชญาพร สิงห์นวล[*1] และ จิราวัลย์ จิตรถเวช[2]
สถาบันบัณฑิตพัฒนบริหารศาสตร์ แขวงคลองจั่น เขตบางกะปิ กรุงเทพฯ 10240

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของตัวประมาณค่าเฉลี่ยของประชากรที่ใช้การสุ่มตัวอย่างแบบง่าย ซึ่งการศึกษานี้ใช้ตัวประมาณ 7 ตัว ได้แก่ ตัวประมาณค่าแบบอัตราส่วน ตัวประมาณค่าแบบการถดถอยเชิงเส้น 2 วิธีที่ใช้ตัวประมาณค่าของซิลส์มาปรับปรุง ตัวประมาณค่าของซิลส์แบบดั้งเดิม ตัวประมาณค่าของซิโซเดียและไวว์ดี (Sisodia and Dwivedi) ตัวประมาณค่าของโคยุนคุและคาดิลาร์ (Koyuncu and Kadilar) และ ตัวประมาณค่าของแชบเบียร์และคุปตะ (Shabbir and Gupta) โดยใช้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) เป็นเกณฑ์ในการเปรียบเทียบของตัวประมาณค่า ข้อมูลที่ใช้เป็นข้อมูลค่าเฉลี่ยจำนวนอาจารย์ต่อสถานศึกษาในเขตกรุงเทพมหานครและเขตปริมณฑลจำนวน 1,009 โรงเรียน ผลการศึกษาพบว่าตัวประมาณค่าเฉลี่ยโดยใช้วิธีการถดถอยเชิงเส้นทั้ง 2 ตัว ที่ใช้ตัวประมาณค่าของซิลส์มาปรับปรุง มีประสิทธิภาพดีกว่าตัวประมาณค่าแบบอื่น สอดคล้องกับผลงานวิจัยทางทฤษฎีที่ได้ศึกษาไว้แล้ว

คำสำคัญ : การสุ่มตัวอย่างอย่างง่าย / การประมาณค่าแบบเอนเอียง / ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย

* *Corresponding Author : emmy__2929@hotmail.com*

1 *นักศึกษา สาขาสถิติ คณะสถิติประยุกต์*

2 *รองศาสตราจารย์ คณะสถิติประยุกต์*

Efficiency Comparison of Conventional Estimators and Modified Searls Estimators

Wichayaporn Singnual[*1] and Jirawan Jitthavech[2]

National Institute of Development Administration, Klong-Chan, Bangkapi, Bangkok 10240

Abstract

The objective of this research was to compare the mean square errors (MSEs) of the means estimated from simple random sampling by seven estimators: ratio estimator, two linear regression estimators based on the modified Searls estimators as well as conventional estimators proposed by Searls, Sisodia and Dwivedi, Koyuncu and Kadilar and Shabbir and Gupta. The study utilized real-life data on the number of teachers per school of 1,009 schools in Bangkok and its surrounding provinces. It was found that two linear regression estimators based on the modified Searls estimators outperformed the rest. This observation is in agreement with that of the theoretical study in the past.

Keywords : Simple Random Sampling / Biased Estimation / Mean Square Error

* *Corresponding Author : emmy__2929@hotmail.com*

1 *Student, Statistics Program, School of Applied Statistics.*

2 *Associate Professor, School of Applied Statistics.*

1. บทนำ

การสำรวจด้วยตัวอย่าง (Sample Survey) เป็นเครื่องมือที่สำคัญที่ใช้ในการเก็บรวบรวมข้อมูล โดยการแจงนับบางหน่วยของประชากร โดยแต่ละหน่วยของประชากรที่ถูกแจงนับจะเป็นไปโดยสุ่ม การเก็บรวบรวมข้อมูลวิธีนี้เป็นวิธีที่ประหยัดเวลาและงบประมาณ จุดมุ่งหมายที่สำคัญของการสำรวจด้วยตัวอย่าง คือ การอนุมานทางสถิติ ซึ่งเป็นการใช้ข้อเท็จจริงที่เก็บรวบรวมได้จากตัวอย่างมาทำการวิเคราะห์เพื่อหาข้อสรุปเกี่ยวกับประชากร ได้แก่ การประมาณพารามิเตอร์ ซึ่งพารามิเตอร์ (Parameter) ที่สำคัญ ได้แก่ ค่าเฉลี่ย ค่ารวมหรือยอดรวม สัดส่วน อัตราส่วนของประชากร

ในการประมาณพารามิเตอร์ตัวหนึ่ง อาจใช้ตัวประมาณที่ต่างกันได้ ซึ่งก็เป็นสิ่งที่จะต้องพิจารณาว่าข้อมูลที่สุ่มมาเป็นตัวอย่างที่นำมาศึกษาเหมาะสมกับตัวประมาณแบบใดมากที่สุด ซึ่งผลจากการประมาณค่าทางสถิติจะมีความถูกต้องและน่าเชื่อถือเพียงใดนั้นย่อมขึ้นอยู่กับวิธีการเลือกตัวอย่าง (Sampling Techniques) เช่น การสุ่มแบบง่าย (Simple Random Sampling) หรือการสุ่มตัวอย่างแบบมีระบบ (Systematic Sampling) เป็นต้น

การเลือกพิจารณาตัวประมาณที่เหมาะสมมีหลักเกณฑ์ในการเลือก โดยสมบัติที่สำคัญ 2 ประการของตัวประมาณที่มักใช้ในการสำรวจด้วยตัวอย่าง ได้แก่ ความไม่เอนเอียงและความคลาดเคลื่อนต่ำ

ในการเปรียบเทียบตัวประมาณ $\hat{\theta}_1$ และ $\hat{\theta}_2$ ของพารามิเตอร์ θ จะพิจารณาคุณสมบัติของ $\hat{\theta}_1$ และ $\hat{\theta}_2$ ในด้านต่างๆ ดังนี้คือ

ความแม่น (Accuracy) ในกรณีที่ตัวประมาณ $\hat{\theta}_1$ และ $\hat{\theta}_2$ ของพารามิเตอร์ θ เป็นตัวประมาณเอนเอียง (Biased) ในการเปรียบเทียบจะพิจารณาที่ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean square error: MSE) เป็นค่าที่ใช้วัดความแม่น (Accuracy) ของตัวประมาณค่าหรือค่าที่ใช้วัดความแปรผัน (Variation) ของตัวประมาณค่าที่เบี่ยงเบนไปจากค่าที่แท้จริงหรือพารามิเตอร์ของประชากร หากตัวประมาณค่าใดให้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุดจะถือว่าตัวประมาณค่านั้นมีประสิทธิภาพสูงสุด

ความเที่ยง (Precision) ในกรณีที่ตัวประมาณค่าเป็นตัวประมาณค่าไม่เอนเอียง (Unbiasness) ซึ่งหมายถึง ถ้า $\hat{\theta}$ เป็นตัวประมาณค่าที่ไม่เอนเอียง (Unbiased Estimator) ของพารามิเตอร์ θ แล้ว $E(\hat{\theta}) = \theta$ การเปรียบเทียบประสิทธิภาพของตัวประมาณจะใช้ความแปรปรวนของตัวประมาณเป็นเครื่องมือวัดความเที่ยงของตัวประมาณนั้น โดยมีเกณฑ์ที่พิจารณาว่าตัวประมาณที่มีความแปรปรวนต่ำที่สุดเป็นตัวประมาณที่ดีที่สุด (Best Estimator) [1-4]

เนื่องจากปัจจุบันมีนักวิจัยหลายท่านได้มีการพัฒนาและนำเสนอวิธีการประมาณค่าวิธีการใหม่ๆ ขึ้นมาเพื่อให้ได้ตัวประมาณค่าที่มีประสิทธิภาพและมีความคลาดเคลื่อนต่ำ ซึ่งได้มีผู้เสนอตัวประมาณค่า ดังนี้

Jitthavech และ Lorchirachoonkul [5] ได้เสนอตัวประมาณค่าเฉลี่ยโดยใช้วิธีอัตราส่วนที่ปรับปรุงใหม่ (New ratio estimator) และ วิธีประมาณค่าเฉลี่ยโดยใช้การถดถอยเชิงเส้นที่ปรับปรุงใหม่ (New linear regression estimators) ผลการศึกษาพบว่า ตัวประมาณค่าเฉลี่ยโดยใช้การถดถอยเชิงเส้นที่ได้นำเสนอไว้ทั้ง 2 วิธี ที่มีการปรับค่าสัมประสิทธิ์การถดถอยที่ใช้ในการคำนวณตัวประมาณค่าใช้สารสนเทศจาก x และ y และใช้สารสนเทศเฉพาะ x พบว่า ตัวประมาณค่าทั้ง 2 นี้มีประสิทธิภาพดีกว่าวิธีประมาณค่าเฉลี่ยของซิลส์แบบดั้งเดิม วิธีประมาณค่าของซิลส์ที่ใช้อัตราส่วนปรับปรุงใหม่ ตัวประมาณค่าของซิโซเดียและไวว์ดี (Sisodia and Dwivedi)

Boonrod [6] ได้ศึกษาเปรียบเทียบประสิทธิภาพตัวประมาณค่าสำหรับค่าเฉลี่ยประชากร ในการเลือกตัวอย่างแบบง่ายในการเลือกตัวอย่างแบบชั้นภูมิ ระหว่างตัวประมาณค่าอัตราส่วนแบบแยก (Separate) ของโคยุนคุและคาดิลาร์ (Koyuncu and Kadilar) และตัวประมาณค่าอัตราส่วนผลคูณแบบแยกของวิชวากาและซิงส์ (Vishwakarma and Singh) โดยใช้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณค่าในการเปรียบเทียบ พบว่า ตัวประมาณค่าอัตราส่วนแบบแยกของ โคยุนคุและคาดิลาร์ มีประสิทธิภาพดีกว่าตัวประมาณค่าอัตราส่วนผลคูณแบบแยกของ วิชวากาและซิงส์ ในทุกกรณี ยกเว้น $\rho^2_{XYh} = 1$ และผลการคำนวณเชิงตัวเลขสอดคล้องกับการศึกษาเชิงทฤษฎี

Tantrakul [7] ได้ศึกษาเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์ในตัวแบบการถดถอยเชิงเส้นแบบพหุ โดยใช้

ตัวแปรอิสระ 2 ตัวแปร และทำการศึกษาวิธีการประมาณ 3 วิธี คือ วิธีกำลังสองน้อยที่สุด วิธีตัวประมาณ LTS และวิธีตัวประมาณ GM เกณฑ์ที่ใช้ในการเปรียบเทียบ คือ ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (MSE) กระทำภายใต้เงื่อนไขของข้อมูลที่มีค่าผิดปกติระดับไม่รุนแรงปะปนอยู่ในตัวแปรอิสระและตัวแปรตาม จำลองข้อมูลด้วยเทคนิคมอนติคาร์โล กระทำซ้ำ 1,000 ครั้ง ในแต่ละกรณี ผลการวิจัย โดยทั่วไปวิธีตัวประมาณ GM ให้ค่า MSE ต่ำที่สุด แต่บางสถานการณ์วิธีกำลังสองน้อยที่สุด และวิธีตัวประมาณ GM ให้ค่า MSE ใกล้เคียงกัน นอกจากนี้ ค่า MSE จะแปรผกผันกับปัจจัยต่อไปนี้ ได้แก่ องศาความเป็นอิสระ ขนาดตัวอย่าง และอัตราส่วนค่าผิดปกติของตัวแปรอิสระ

Chaloeiwaret [8] ได้ศึกษาเปรียบเทียบวิธีสำหรับการประมาณค่าอัตราส่วนประชากร โดยการใช้ (1) อัตราส่วนระหว่างค่าเฉลี่ยตัวอย่างของตัวแปรเป้าหมายและตัวแปรช่วย และ (2) ค่าเฉลี่ยของอัตราส่วนระหว่างค่าของตัวแปรเป้าหมายและค่าของตัวแปรช่วยที่ได้จากตัวอย่าง ภายใต้การแจกแจงปกติสองตัวแปร การแจกแจงปัวซงล็อกนอร์มัลสองตัวแปร และการแจกแจงโคชีสองตัวแปร ผลการศึกษาพบว่า ประชากรมีลักษณะการแจกแจงปกติสองตัวแปร และการแจกแจงปัวซงล็อกนอร์มัลสองตัวแปร เมื่อมีขนาดตัวอย่างเล็ก ค่าความน่าจะเป็นครอบคลุมของทั้ง 2 วิธี มีค่าใกล้เคียงกันทุกกรณี แต่เมื่อขนาดตัวอย่างใหญ่ขึ้นการใช้อัตราส่วนของค่าเฉลี่ยจะให้ค่าความน่าจะเป็นครอบคลุมที่สูงกว่าการใช้ค่าเฉลี่ยของอัตราส่วน ส่วนประชากรมีลักษณะการแจกแจงโคชีสองตัวแปร การใช้อัตราส่วนของค่าเฉลี่ยจะให้ค่าความน่าจะเป็นครอบคลุมที่สูงกว่าการใช้ค่าเฉลี่ยของอัตราส่วนทุกกรณี

Searls [9] ได้เสนอตัวประมาณของประชากรค่าเฉลี่ยแบบเอนเอียง เมื่อทราบค่าสัมประสิทธิ์การผันแปร และพบว่า ค่าเฉลี่ยตัวอย่างแบบเอนเอียงมีค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) ต่ำกว่าค่าเฉลี่ยแบบดั้งเดิม

Sisodia และ Dwivedi [10] ได้เสนอตัวประมาณค่าที่ใช้ค่าสัมประสิทธิ์การผันแปรของตัวแปรช่วย สำหรับการประมาณค่าเฉลี่ยประชากร เมื่อตัวแปรศึกษาและตัวแปรช่วยมีความสัมพันธ์กันสูง พบว่า ตัวประมาณค่าที่ได้นำเสนอมีประสิทธิภาพดีกว่าตัวประมาณค่าอัตราส่วนแบบเดิม

Koyuncu และ Kadilar [11] ได้ทำการปรับปรุงวิธีการประมาณค่าเฉลี่ยประชากรโดยได้เสนอตัวประมาณค่าเฉลี่ยโดยใช้อัตราส่วน พบว่า ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของวิธีที่นำเสนอมีค่าน้อยกว่า

Shabbir และ Gupta [12] ได้เสนอวิธีประมาณค่าอัตราส่วนแบบเลขชี้กำลัง (exponential)

งานวิจัยนี้ใช้ข้อมูลจริงเพื่อเปรียบเทียบประสิทธิภาพของตัวประมาณ 7 ตัวโดยใช้วิธีต่างๆ ดังนี้ 1) วิธีประมาณค่าของซิลส์ (Searl : S) 2) วิธีประมาณค่าของซิโซเดียและไวว์ดี (Sisodia and Dwivedi :SD) 3) วิธีประมาณค่าของโคยุนคุและคาดิลาร์ (Koyuncu and Kadilar : KK) 4) วิธีประมาณค่าของแชบเบียร์และคุปตะ (Shabbir and Gupta : SG) 5) วิธีประมาณค่าเฉลี่ยโดยใช้อัตราส่วนที่ปรับปรุงใหม่ (New ratio estimator : rwyx) และ 6) วิธีประมาณค่าเฉลี่ยโดยใช้การถดถอยเชิงเส้นที่ปรับปรุงใหม่ ซึ่งมี 2 ตัวที่ใช้น้ำหนักแตกต่างกัน (New linear regression estimators : lrwyx, lrwx)

2. วิธีดำเนินการวิจัย

2.1 ขนาดตัวอย่างและการสุ่มตัวอย่าง

หาขนาดตัวอย่างที่เหมาะสมสำหรับใช้ในการสุ่มตัวอย่างจากประชากรทั้งหมด จำนวน 1,009 โรงเรียน โดยคำนวณจากสูตรของ Taro Yamane [11]

$$n = \frac{N}{1+Ne^2} \quad \textbf{(1)}$$

เมื่อ n = ขนาดตัวอย่าง
N = ขนาดประชากร
e = ความคลาดเคลื่อนของการสุ่มตัวอย่างที่ยอมรับได้

ในการวิจัยครั้งนี้กำหนดค่าความคลาดเคลื่อนเป็น 0.05 ซึ่งได้ขนาดตัวอย่างเท่ากับ 187 หน่วย แล้วทำการสุ่มตัวอย่างด้วยวิธีการสุ่มตัวอย่างแบบง่ายไม่คืนที่ โดยทำการสุ่มตัวอย่างจำนวน 30 ครั้งในแต่ละวิธี

2.2 การประมาณค่าเฉลี่ย

การประมาณค่าเฉลี่ยและคำนวณค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของข้อมูลด้วยวิธี ดังนี้

2.2.1 วิธีประมาณค่าเฉลี่ยของซิลส์ [9]

ในการสุ่มตัวอย่างแบบง่ายจะมีขนาดของประชากรเป็น N และขนาดตัวอย่างเป็น n ซึ่ง Searls [9] ได้เสนอค่าเฉลี่ยตัวอย่างแบบเอนเอียงซึ่งมีค่าความคลาดเคลื่อนกำลังสองเฉลี่ยน้อยกว่าค่าเฉลี่ยวิธีดั้งเดิม $\bar{y}$ และทราบค่าสัมประสิทธิ์การผันแปร C_y และ $\gamma = 1/n - 1/N$ ดังนั้นตัวประมาณค่าเฉลี่ยประชากรที่เอนเอียงของซิลส์ เป็นดังนี้

$$\bar{y}'_S = \frac{\bar{y}}{1+\gamma C_y^2} \quad \textbf{(2)}$$

ค่าเอนเอียงของตัวประมาณค่าเฉลี่ยประชากร (Bias) และค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) ของตัวประมาณค่าของซิลส์ เป็นดังนี้

$$\mathrm{Bias}\left(\bar{y}'_S\right) = \frac{\gamma C_y^2 \bar{Y}}{1+C_y^2} \quad \textbf{(3)}$$

$$\mathrm{MSE}\left(\bar{y}'_S\right) = \frac{\gamma C_y^2 \bar{Y}^2}{1+\gamma C_y^2} \quad \textbf{(4)}$$

2.2.2 วิธีประมาณค่าเฉลี่ยของซิโซเดียและไวว์ดี [10]

นักวิจัยหลายท่านได้เสนอตัวประมาณค่าที่ใช้ค่าสัมประสิทธิ์การผันแปร C_x ของตัวแปรช่วย (X) สำหรับการประมาณค่าเฉลี่ยประชากร เมื่อตัวแปรศึกษาและตัวแปรช่วยมีความสัมพันธ์กันสูง ซึ่งตัวประมาณค่าเฉลี่ยแบบไม่เอนเอียงของซิโซเดียและไวว์ดี เป็นดังนี้

$$\bar{y}_{SD} = \left(\frac{\bar{X}+C_x}{\bar{x}+C_x}\right) \quad \textbf{(5)}$$

ค่าเอนเอียงของตัวประมาณค่าเฉลี่ยประชากร (Bias) และค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) ของตัวประมาณค่าซิโซเดียและไวว์ เป็นดังนี้

$$\mathrm{Bias}\left(\bar{y}_{SD}\right) \approx \gamma \bar{Y} \frac{\bar{X}}{\bar{X}+C_x}\left(\frac{\bar{X}C_x^2}{\bar{X}+C_x} - \rho C_y C_x\right) \quad \textbf{(6)}$$

$$\mathrm{MSE}\left(\bar{y}_{SD}\right) \approx \gamma \bar{Y}^2 \left[C_y^2 + \left(\frac{\bar{X}}{\bar{X}+C_x}\right)\left(\frac{\bar{X}C_x^2}{\bar{X}+C_x} - 2\rho C_y C_x\right)\right], \quad \textbf{(7)}$$

2.2.3 วิธีประมาณค่าเฉลี่ยของโคยุนคุและคาดิลาร์ [11]

Gupta และ Shabbir [14] ได้ทำการปรับปรุงวิธีการประมาณค่าเฉลี่ยประชากรโดยได้เสนอตัวประมาณแบบอัตราส่วน เป็นดังนี้

$$\bar{y}_p = \left[w_1 \bar{y} + w_2\left(\bar{X}+\bar{x}\right)\right]\left(\frac{\eta \bar{X}+\lambda}{\eta \bar{x}+\lambda}\right) \quad \textbf{(8)}$$

เมื่อ η และ λ เป็นค่าคงที่หรือฟังก์ชันที่ทราบค่าพารามิเตอร์ของตัวแปรช่วยของ w_1 และ w_2 เป็นค่าคงที่ สำหรับค่าเอนเอียงของตัวประมาณค่าเฉลี่ยประชากร (Bias) และค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) ของตัวประมาณค่าเฉลี่ยของ $\bar{y}_p$ เป็นดังนี้

$$\mathrm{Bias}\left(\bar{y}_p\right) \approx \left(w_1 - 1\right)\bar{Y} + \gamma\left[w_1 \bar{Y}\left(\tau^2 C_x^2 - \tau \rho C_y C_x\right) + w_2 \bar{X} \tau C_x^2\right], \quad \textbf{(9)}$$

$$\bar{y}_p = \left[w_1 \bar{y} + w_2\left(\bar{X}+\bar{x}\right)\right]\left(\frac{\eta \bar{X}+\lambda}{\eta \bar{x}+\lambda}\right) \quad \textbf{(10)}$$

เมื่อ $\mathrm{MSE}\left(\bar{y}_p\right) \approx \gamma \bar{Y}^2 C_y^2\left(1-\rho^2\right)$ (Cochran, 1977) และ $\tau = \frac{\eta \bar{X}}{\eta \bar{X}+\lambda} < 1$ ค่าของ w_1 และ w_2 ที่ทำให้ $\mathrm{MSE}\left(\bar{y}_p\right)$ มีค่าต่ำสุด คือ

$$w_1 = \frac{1-\gamma C_x^2 \tau^2}{1+\gamma\left[C_y^2\left(1-\rho^2\right) - \tau^2 C_x^2\right]} \quad \textbf{(11)}$$

$$w_2 = \frac{\bar{Y}}{\bar{X}}\left\{\tau + \frac{\left(1-\gamma C_x^2\tau^2\right)\left(\rho C_y - 2\tau C_x\right)}{C_x\left[1+\gamma C_y\left(1-\rho^2\right)-\gamma C_x^2\tau^2\right]}\right\}, \quad \textbf{(12)}$$

และกำหนดค่า η และ λ แสดงดังตารางที่ 1

ตารางที่ 1 ค่า η และ λ

KK Estimator	η	λ
$\bar{y}_{p(0)}$	0	1
$\bar{y}_{p(1)}$	1	0
$\bar{y}_{p(2)}$	1	C_x
$\bar{y}_{p(3)}$	1	$\beta_{2(x)}$
$\bar{y}_{p(4)}$	$\beta_{2(x)}$	C_x
$\bar{y}_{p(5)}$	C_x	$\beta_{2(x)}$

2.2.4 วิธีประมาณค่าเฉลี่ยของแชบเบียร์และคุปตะ [10] ได้เสนอวิธีประมาณค่าอัตราส่วนแบบเลขชี้กำลัง (exponential) เป็นดังนี้

$$\bar{y}_{SG} = k_1\bar{y} + k_2\left(\bar{X}-\bar{x}\right)\exp\left[\frac{\bar{X}-\bar{x}}{\left(2N+1\right)\bar{X}+\bar{x}}\right], \quad \textbf{(13)}$$

เมื่อ

$$k_1 = \frac{1-\frac{\gamma C_x^2}{8\left(1+N\right)^2}}{1+\gamma C_y^2\left(1-\rho^2\right)}$$

$$k_2 = \frac{\bar{Y}}{\bar{X}}\left[\frac{1}{2\left(1+N\right)} - k_1\left(\frac{1}{1+N}-\rho\frac{C_y}{C_x}\right)\right]$$

ค่าเอนเอียงของตัวประมาณค่าเฉลี่ยประชากร (Bias) และค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) ของ $\bar{y}_{SG}$ เป็นดังนี้

$$\mathrm{Bias}\left(\bar{y}_{SG}\right) \approx \bar{Y}\left\{k_1 - 1 + k_1\gamma\left[\frac{3C_x^2}{8\left(1+N\right)^2} - \frac{\rho C_x C_y}{2\left(1+N\right)}\right]\right\} + \frac{\gamma C_x^2 k_2\bar{X}}{2\left(1+N\right)^2} \quad \textbf{(14)}$$

$$\mathrm{MSE}\left(\bar{y}_{SG}\right) \approx \bar{Y}^2\left\{1 - \frac{\gamma C_x^2}{4\left(1+N\right)^2} - \frac{\left[1-\frac{\gamma C_x^2}{8\left(1+N\right)^2}\right]^2}{1+\gamma C_y^2\left(1-\rho^2\right)}\right\}, \quad \textbf{(15)}$$

2.2.5 วิธีประมาณค่าเฉลี่ยโดยใช้อัตราส่วนที่ปรับปรุงใหม่ เป็นวิธีที่ได้มีผู้นำเสนอไว้แล้ว [5] เมื่อตัวแปรศึกษา (Y) มีความสัมพันธ์กันสูงกับตัวแปรช่วย (X) การประมาณค่าอัตราส่วนจะใช้สำหรับประมาณค่าเฉลี่ยประชากร ซึ่งตัวประมาณอัตราส่วนแบบใหม่จะใช้สัมประสิทธิ์การผันแปรของตัวแปร Y และตัวแปร X ดังนี้

$$\bar{y}'_{rwyx} = \frac{\bar{X}}{\bar{x}'_{wx}}\bar{y}'_{wy} \quad \textbf{(16)}$$

เมื่อ $\bar{y}'_{wy} = \bar{y}\Big/\left(1+\gamma C_y^2\right)$, $\bar{x}' = \bar{x}\Big/\left(1+\gamma C_x^2\right)$ และ $\bar{X}$ คือ ค่าเฉลี่ยประชากรของตัวแปรช่วย และกรณีทราบค่าสัมประสิทธิ์การผันแปร จะได้ตัวประมาณค่าแบบอัตรา ส่วนแบบใหม่ คือ C_x, C_y

$$\bar{y}'_{rwyx} = \frac{1+\gamma C_x^2}{1+\gamma C_y^2}\bar{y}_r \quad \textbf{(17)}$$

ค่าเอนเอียงของตัวประมาณค่าเฉลี่ยประชากร (Bias) และค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) ของ $\bar{y}'_{rwyx}$ เป็นดังนี้

$$\mathrm{Bias}\left(\bar{y}'_{rwyx}\right) = \frac{1+\gamma C_x^2}{1+\gamma C_y^2}\mathrm{Bias}\left(\bar{y}_r\right) - \frac{\gamma\left(C_y^2-C_x^2\right)}{1+\gamma C_y^2}\bar{Y}, \quad \textbf{(18)}$$

$$\mathrm{MSE}\left(\bar{y}'_{rwyx}\right) = \left(\frac{1+\gamma C_x^2}{1+\gamma C_y^2}\right)\mathrm{MSE}\left(\bar{y}_r\right) \quad \textbf{(19)}$$

2.2.6 วิธีประมาณค่าเฉลี่ยโดยใช้การถดถอยเชิงเส้นที่ปรับปรุงใหม่ [5]

ตามที่ได้มีการกำหนดตัวประมาณค่าการถดถอยเชิงเส้นแบบเดิมไว้แล้วนั้น Cochran [15] จะได้ตัวประมาณค่าเฉลี่ยโดยใช้การถดถอยเชิงเส้นที่ปรับปรุงใหม่ วิธีที่ 1 เป็นดังนี้

$$\bar{y}'_{lrwyx} = \bar{y}'_{wy} + b_{wyx}\left(\bar{X} - \bar{x}'_{wx}\right) \quad \textbf{(20)}$$

ซึ่งค่าของ b_{wyx} ที่จะทำให้ความแปรปรวนของ $\bar{y}'_{rwyx}$ มีค่าต่ำสุดคือ

$$b_{wyx} = \frac{1+\gamma C_x^2}{1+\gamma C_y^2} b \quad \textbf{(21)}$$

เมื่อ $b = \rho S_y / S_X$ (Cochran, 1977) โดยจะแทนค่า b_{wyx} จากสมการที่ 25 เข้าไปในสมการที่ 24 และค่าเอนเอียงของตัวประมาณค่าเฉลี่ยประชากร (Bias) ของ $\bar{y}'_{rwyx}$ เป็นดังนี

$$\mathrm{Bias}\left(\bar{y}'_{lrwyx}\right) = \mathrm{Bias}\left(\bar{y}'_{wy}\right)\left(1 - \rho\frac{C_x}{C_y}\right) \quad \textbf{(22)}$$

และค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) ของ $\bar{y}'_{rwyx}$ เป็นดังนี้

$$\mathrm{MSE}\left(\bar{y}'_{lrwyx}\right) \approx \frac{\mathrm{MSE}\left(\bar{y}_{lr}\right)}{\left(1+\gamma C_y^2\right)^2} \quad \textbf{(23)}$$

กรณีถ้ามีเฉพาะ C_x ที่ถูกนำมาใช้ในการประมาณค่า จะได้ตัวประมาณค่าเฉลี่ยโดยใช้การถดถอยเชิงเส้นที่ปรับปรุงใหม่วิธีที่ 2 เป็นดังนี้

$$\bar{y}'_{lrwx} = \bar{y}'_{wx} + b_{wx}\left(\bar{X} - \bar{x}'_{wx}\right) \quad \textbf{(24)}$$

ซึ่งจะได้ว่า $b_{wx} = b$ และค่าเอนเอียงและความคลาดเคลื่อนกำลังสองเฉลี่ย เป็นดังนี้

$$\mathrm{Bias}\left(\bar{y}'_{lrwx}\right) = \mathrm{Bias}\left(\bar{y}'_{wx}\right)\left(1 - \rho\frac{C_y}{C_x}\right) \quad \textbf{(25)}$$

$$\mathrm{MSE}\left(\bar{y}'_{lrwx}\right) \approx \frac{\mathrm{MSE}\left(\bar{y}_{lr}\right)}{\left(1+\gamma C_x^2\right)^2} \quad \textbf{(26)}$$

3. วิธีการศึกษา

ในการเก็บรวบรวมข้อมูล ใช้ข้อมูลจริงจากจำนวนสถานศึกษาในเขตกรุงเทพมหานครและเขตปริมณฑล จำนวน 1,009 โรงเรียน โดยกำหนดให้ตัวแปรศึกษา (Y) แทนจำนวนอาจารย์ และตัวแปรช่วย (X) แทนจำนวนนักเรียน [16] แล้วทำการสุ่มตัวอย่างจากข้อมูลจริง โดยวิธีการสุ่มตัวอย่างแบบง่าย มาจำนวน 187 หน่วย ด้วย Program SAS และกระทำการสุ่มซ้ำจำนวน 30 ครั้ง ในแต่ละวิธี แบบไม่คืนที่ เพื่อประมาณค่าเฉลี่ยของจำนวนอาจารย์ต่อสถานศึกษาในเขตกรุงเทพมหานครและเขตปริมณฑล แล้วนำข้อมูลที่ได้จากการสุ่มตัวอย่างหาค่าความคลาดเคลื่อนกำลังสองเฉลี่ย และนำค่าความคลาดเคลื่อนกำลังสองเฉลี่ยที่ได้ในแต่ละวิธีมาเปรียบเทียบกัน โดยวิธีประมาณค่าเฉลี่ยที่ให้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุด มีประสิทธิภาพสูงสุด

4. ผลการศึกษา

การศึกษาในครั้งนี้ได้เปรียบเทียบประสิทธิภาพของตัวประมาณค่าเฉลี่ยในการสุ่มตัวอย่างแบบง่าย โดยวิธีการประมาณที่มีผู้เสนอแนะไว้ 6 วิธี คือ 1) วิธีประมาณค่าของซิลส์ 2) วิธีประมาณค่าของซิโซเดียและไวว์ดี 3) วิธีประมาณค่าของโคยุนคุและคาดิลาร์ 4) วิธีประมาณค่าของแชบเบียร์และคุปตะ 5) วิธีประมาณค่าเฉลี่ยโดยใช้อัตราส่วนที่ปรับปรุงใหม่ และ 6) วิธีประมาณค่าเฉลี่ยโดยใช้การถดถอยเชิงเส้นที่ปรับปรุงใหม่ โดยใช้ข้อมูลสถานศึกษาในเขตกรุงเทพมหานครและเขตปริมณฑล จำนวน 1,009 โรงเรียน โดยมีค่าพารามิเตอร์ประชากร ที่คำนวณได้แสดงดังตารางที่ 2

ตารางที่ 2 พารามิเตอร์ประชากร

	μ	σ^2	C.V.
X	429.18930	184234.96000	1.00009
Y	25.81368	565.25295	0.92103

สำหรับค่าสถิติที่คำนวณได้จากข้อมูลตัวอย่างที่สุ่มมาได้จำนวน 187 หน่วย จากประชากรทั้งหมด โดยกระทำการสุ่มซ้ำจำนวน 30 ครั้งในแต่ละวิธี แล้วนำมาหาค่าเฉลี่ยในแต่ละวิธี ซึ่งค่าสถิติต่างๆ อ้างอิงตามตารางที่ 3

ตารางที่ 3 ค่าสถิติจากข้อมูลตัวอย่างจำนวน 30 ครั้ง

ครั้งที่	$\bar{x}$	$\bar{y}$	C_x	C_y	s_x^2	s_y^2
1	417.74866	25.54011	0.97054	0.87927	164381.48	504.30349
2	475.35294	28.56684	1.03724	1.04727	243102.30	895.0317981
3	435.62032	24.25134	0.98019	0.87938	182321.83	454.8020815
4	474.28342	28.73262	1.07824	1.09258	261522.34	985.5087689
5	418.57219	26.35294	0.97935	0.85298	168042.63	505.2833650
6	457.99465	26.44920	1.00448	0.88409	211643.89	546.7863838
7	427.16578	25.78075	1.01272	0.98471	187141.78	644.4839285
8	382.72727	23.35294	0.93858	0.79671	129038.19	346.1650854
9	414.94118	24.13369	0.90689	0.82084	141606.84	392.4282675
10	384.37968	23.08556	1.04748	0.89908	162111.53	430.7990915
11	389.79144	24.16043	0.99001	0.85972	148916.56	431.4472428
12	379.73262	22.70588	0.95210	0.86242	130714.46	383.4560405
13	420.44920	24.86096	1.04365	0.90193	192548.38	502.7870163
14	445.24599	26.51337	1.07421	0.94622	228761.09	629.3801967
15	490.35829	28.98396	0.93185	0.84159	208794.24	594.9943649
16	435.90909	26.39037	1.15086	1.09231	251675.18	830.9596918
17	430.60428	25.51337	0.81860	0.74329	124250.83	359.6275085
18	450.34759	27.33155	0.98315	0.89852	196036.20	603.0937841
19	467.80749	27.51337	1.10598	1.04990	267687.58	834.4124547
20	419.58289	25.02674	0.91586	0.84293	147670.24	445.0369156
21	429.30481	25.83422	1.01478	0.87199	189792.16	507.4723708
22	428.40107	25.65241	0.85619	0.82738	134535.52	450.4753033
23	439.27807	25.33155	0.99948	0.93698	192763.67	563.3626013
24	435.27273	26.13369	1.00175	0.87491	190127.13	522.7938589
25	391.43850	24.49198	1.04396	0.96290	166992.20	556.1760106
26	394.25134	24.69519	0.97131	0.89117	146643.15	484.3313208
27	376.76471	22.43850	0.99248	0.82679	139825.43	344.1722730
28	377.21390	24.01070	0.94918	0.85048	128196.42	416.9998850
29	413.21390	24.43850	0.95119	0.88386	154484.51	466.5701225
30	455.85027	26.95187	1.05615	0.85646	231790.98	532.8310045
ค่าเฉลี่ย	425.32014	25.50749	0.99195	0.89862	180770.62467	538.86574

ในการเปรียบเทียบประสิทธิภาพของตัวประมาณค่าผู้วิจัยได้ทำการเปรียบเทียบประสิทธิภาพของตัวประมาณค่า $\bar{y}'_S$, $\bar{y}_{SD}$, $\bar{y}_p$, $\bar{y}_{SG}$, $\bar{y}'_{rwyx}$, $\bar{y}'_{lrwyx}$ และ $\bar{y}'_{lrwx}$ โดยใช้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยเป็นเกณฑ์ในการเปรียบเทียบ ซึ่งวิธีที่ให้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุด จะเป็นตัวประมาณค่าที่ดีที่สุด ซึ่งผลการเปรียบเทียบแสดงได้ดังตารางที่ 4

ตารางที่ 4 ค่าเอนเอียงและค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณค่า

Estimators	Bias	MSE
$\bar{y}'_{lrwx}$	-0.01639	0.44267
$\bar{y}'_{lrwyx}$	-0.00087	0.44330
$\bar{y}_{p(0)}$	-0.01729	0.44622
$\bar{y}_{p(1)}$	-0.01729	0.44622
$\bar{y}_{p(2)}$	-0.01729	0.44622
$\bar{y}_{p(3)}$	-0.01729	0.44622
$\bar{y}_{p(4)}$	-0.01729	0.44622
$\bar{y}_{p(5)}$	-0.01729	0.44622
$\bar{y}_{SG}$	-0.01733	0.44622
$\bar{y}_{SD}$	0.02018	0.54887
$\bar{y}'_{rwyx}$	0.04005	0.55214
$\bar{y}'_S$	-0.09124	2.35528

จากตารางที่ 4 พบว่า ตัวประมาณค่าที่ให้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุด คือ ตัวประมาณค่าเฉลี่ยโดยใช้การถดถอยเชิงเส้นที่ปรับปรุงใหม่ คือ $\bar{y}'_{lrwx}$ ซึ่งมีความคลาดเคลื่อนกำลังสองเฉลี่ยเท่ากับ 0.44267

5. สรุป อภิปรายผล และข้อเสนอแนะ

จากผลการเปรียบเทียบประสิทธิภาพตัวประมาณค่าเฉลี่ยทั้ง 6 วิธี คือ วิธีประมาณค่าของซิลส์ วิธีประมาณค่าของซิโซเดียและไวว์ดี วิธีประมาณค่าของโคยุนคุและคาดิลาร์ วิธีประมาณค่าของแชบเบียร์และคุปตะ วิธีประมาณค่าเฉลี่ยโดยใช้อัตราส่วนที่ปรับปรุงใหม่ และวิธีประมาณค่าเฉลี่ยโดยใช้การถดถอยเชิงเส้นที่ปรับปรุงใหม่ พบว่า ตัวประมาณค่าเฉลี่ยโดยใช้การถดถอยเชิงเส้นที่ปรับปรุงใหม่ทั้ง 2 วิธี คือ $\bar{y}'_{lrwx}$ และ $\bar{y}'_{lrwyx}$ ให้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย คือ 0.44267 และ 0.44330 และค่าเอนเอียง คือ -0.01639 และ -0.00087 ต่ำสุด ซึ่งมีประสิทธิภาพดีกว่าวิธีประมาณค่าแบบอื่นทั้ง 5 วิธี

จากสรุปผลการวิจัย พบว่าตัวประมาณค่าเฉลี่ยโดยใช้การถดถอยเชิงเส้นที่ปรับปรุงใหม่ทั้ง 2 วิธี คือ ที่มีการปรับค่าสัมประสิทธิ์การถดถอยที่ใช้ในการคำนวณตัวประมาณค่าใช้สารสนเทศจาก x และ y และใช้สารสนเทศเฉพาะ x $\bar{y}'_{lrwx}$ และ $\bar{y}'_{lrwyx}$ มีประสิทธิภาพดีกว่าวิธีประมาณค่าของซิลส์ วิธีประมาณค่าของซิโซเดียและไวว์ดี วิธีประมาณค่าของโคยุนคุและคาดิลาร์ วิธีประมาณค่าของแชบเบียร์และคุปตะ และวิธีประมาณค่าเฉลี่ยโดยใช้อัตราส่วนที่ปรับปรุงใหม่และวิธีประมาณค่าเฉลี่ยตัวอย่างเอนเอียงที่ปรับปรุงใหม่ ซึ่งสอดคล้องกับผลการศึกษาของจิราวัลย์ จิตรถเวช และ วิชิต หล่อจีระชุณห์กุล [15]

6. กิตติกรรมประกาศ

ขอขอบคุณ รศ. ดร. วิชิต หล่อจีระชุณห์กุล ที่ให้ข้อคิดเห็นและข้อเสนอแนะต่างๆ ทำให้งานชิ้นนี้สำเร็จด้วยดี

7. เอกสารอ้างอิง

1. Suwattee, S., 2009, Sample Survey : Sampling and Analysis, Office Printing Business Relief Organizations Veteran Nationals, Bangkok. (In Thai).

2. Suwattee, S., 2011, Sampling Theory, Office Printing Business Relief Organizations Veteran Nationals, Bangkok. (In Thai).

3. Suwattee, S., 2010, Inference Statistics Theory, Office Printing Business Relief Organizations Veteran nationals, Bangkok. (In Thai).

4. Suwattee, S. and Team., 2012, Methodology Statistics, Soda Advertising, Bangkok. (In Thai).

5. Jitthavech, J. and Lorchirachoonkul, V., 2013, "Estimators in simple Random sampling: Searls approach," *Sonklanakarin Journal Science*, 35, pp. 749-760.

6. Boonrod, T., 2012, Efficiency Comparisons of Separate Estimators in Stratified Random Sampling, Master of Science Statistics Thesis, Burapa University. (In Thai).

7. Tantrakul, O., 2012, Comparison of Robust Regression Methods in Multiple Linear Regression Model, Master of Science Statistics Thesis, Khon Kaen University. (In Thai).

8. Chaloeiwaret, M., 2010, A Comparative Study on Population Ratio Estimation in Simple Random Sampling, Master of Science Applied Statistics Thesis, Thammasat University. (In Thai).

9. Searls, D.T., 1964, "The Utilization of a Known Coefficient of Variation in the Estimation Procedure," *Journal of the American Statistical Association*, 59, pp. 1225-1226.

10. Sisodia, B.V.S. and Dwivedi, V.K., 1981, "A Modified Ratio Estimator using Known Coefficient of Variation of Auxiliary Variable," *Journal of the Indian Society of Agricultural Statistics*, 33, pp. 13-18.

11. Koyuncu, N. and Kadilar, C., 2010, "On Improvement in Estimating Population Mean in Stratified Random Sampling," *Journal of Applied Statistics*, 37, pp. 999-1013.

12. Shabbir, J. and Gupta, S., 2011, "On Estimating Finite Population Mean in Simple and Stratified Sampling," *Communications in Statistics-Theory and Methods*, 40, pp. 199-212.

13. Yamane, T., 1973, Statistics : An Introductory Analysis, Harper and Row Publication, New York.

14. Gupta, S. and Shabbir, J., 2008, "On Improvement in Estimating the Population Mean in Simple Random Sampling," *Journal of Applied Statistics*, 35, pp. 559-566

15. Cochran, W.G., 1977, Sampling Techniques, John Wiley and Sons, New York, U.S.A.

16. http://data.bopp-obec.info/emis/area__school.php [2013, May 26].