ตัวประมาณไม่เอนเอียงโดยการใช้สารสนเทศของตัวแปรช่วยสำหรับค่าเฉลี่ยประชากรจำกัดภายใต้การเลือกตัวอย่างสองเฟส

วิทยานิพนธ์

ของ

อธิปกรณ์ นาถมทอง

เสนอต่อมหาวิทยาลัยมหาสารคาม เพื่อเป็นส่วนหนึ่งของการศึกษาตามหลักสูตร

ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการจัดการสถิติ

พฤศจิกายน 2566

ตัวประมาณไม่เอนเอียงโดยการใช้สารสนเทศของตัวแปรช่วยสำหรับค่าเฉลี่ยประชากรจำกัดภายใต้การเลือกตัวอย่างสองเฟส

วิทยานิพนธ์

ของ

อธิปกรณ์ นาถมทอง

เสนอต่อมหาวิทยาลัยมหาสารคาม เพื่อเป็นส่วนหนึ่งของการศึกษาตามหลักสูตร

ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการจัดการสถิติ

พฤศจิกายน 2566

Unbiased Estimators Using Auxiliary Information for Finite Population Mean Under Two-Phase Sampling

Athipakon Nathomthong

A Thesis Submitted in Partial Fulfillment of Requirements
for Master of Science (Statistical Management Science)
November 2023

คณะกรรมการสอบวิทยานิพนธ์ ได้พิจารณาวิทยานิพนธ์ของนายอธิปกรณ์ นาถมทอง แล้วเห็นสมควรรับเป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการจัดการสถิติ ของมหาวิทยาลัยมหาสารคาม

คณะกรรมการสอบวิทยานิพนธ์

.... ประธานกรรมการ
(ผศ. ดร. พรรณรัตน์ ก้วยเจริญพานิชก์)

.... อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก
(รศ. ดร. นิภาพร ชุติมันต์)

.... กรรมการ
(อ. ดร. สุภาวดี วิชิตชาญ)

.... กรรมการ
(ผศ. ดร. มนชยา เจียงประดิษฐ์)

มหาวิทยาลัยอนุมัติให้รับวิทยานิพนธ์ฉบับนี้ เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญา วิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการจัดการสถิติ ของมหาวิทยาลัยมหาสารคาม

....	....
(ศ. ดร. ไพโรจน์ ประมวล)	(รศ. ดร. กริสน์ ชัยมูล)
คณบดีคณะวิทยาศาสตร์	คณบดีบัณฑิตวิทยาลัย

ชื่อเรื่อง	ตัวประมาณไม่เอนเอียงโดยการใช้สารสนเทศของตัวแปรช่วยสำหรับค่าเฉลี่ยประชากรจำกัดภายใต้การเลือกตัวอย่างสองเฟส
ผู้วิจัย	อธิปกรณ์ นาถมทอง
อาจารย์ที่ปรึกษา	รองศาสตราจารย์ ดร. นิภาพร ชุติมันต์
ปริญญา	วิทยาศาสตรมหาบัณฑิต **สาขาวิชา** วิทยาการจัดการสถิติ
มหาวิทยาลัย	มหาวิทยาลัยมหาสารคาม **ปีที่พิมพ์** 2566

บทคัดย่อ

ในงานวิจัยนี้ ผู้วิจัยศึกษาและพัฒนาทฤษฎีเกี่ยวกับตัวประมาณอัตราส่วนและตัวประมาณผลคูณไม่เอนเอียง ที่เสนอโดย Mahanty & Mishra (2020) โดยการใช้สารสนเทศของตัวแปรช่วยสำหรับค่าเฉลี่ยประชากรจำกัด ภายใต้การเลือกตัวอย่างสองเฟส และเปรียบเทียบประสิทธิภาพของตัวประมาณที่ผู้วิจัยนำเสนอกับตัวประมาณอัตราส่วนและตัวประมาณผลคูณ พบว่า ตัวประมาณที่ผู้วิจัยเสนอ มีประสิทธิภาพสูงกว่าตัวประมาณอัตราส่วนและตัวประมาณผลคูณ ในเชิงทฤษฎีและเชิงตัวเลข

คำสำคัญ : การเลือกตัวอย่างสองเฟส, ตัวประมาณอัตราส่วน, ตัวประมาณผลคูณ, สารสนเทศของตัวแปรช่วย, ตัวประมาณไม่เอนเอียง

TITLE	Unbiased Estimators Using Auxiliary Information for Finite Population Mean Under Two-Phase Sampling		
AUTHOR	Athipakon Nathomthong		
ADVISORS	Associate Professor Nipaporn Chutiman , Ph.D.		
DEGREE	Master of Science	MAJOR	Statistical Management Science
UNIVERSITY	Mahasarakham University	YEAR	2023

ABSTRACT

In this paper, the researcher studies and develops the theory of unbiased ratio estimators and unbiased product estimators based on Mahanty & Mishra (2020) by using auxiliary information for finite population mean under Two-Phase sampling. Comparing the efficiency of the proposed estimator with the general ratio estimator and the general product estimator, it was found that the proposed estimator is more efficient than the general ratio estimator and general product estimator both theoretically and with the numerical illustration.

Keyword : Two-Phase Sampling, Ratio Estimator, Product Estimator, Auxiliary Variable, Unbiased Estimator

กิตติกรรมประกาศ

งานวิจัยฉบับนี้สำเร็จลุล่วงได้ด้วยความกรุณาจาก รองศาสตราจารย์ ดร.นิภาพร ชุติมันต์ อาจารย์ที่ปรึกษา ที่ช่วยเหลือ แนะนำ ถ่ายทอดความรู้ ให้คำปรึกษา ตรวจสอบแก้ไขข้อบกพร่องต่าง ๆ คำชี้แนะในการดำเนินการทุกขั้นตอน ด้วยความเอาใจใส่อย่างดียิ่ง ข้าพเจ้าตระหนักถึงความตั้งใจจริง และความทุ่มเทของอาจารย์และขอกราบขอบพระคุณเป็นอย่างสูงไว้ ณ ที่นี้

ขอขอบพระคุณ ผู้ช่วยศาสตราจารย์ ดร.พรรณรัตน์ ก้วยเจริญพานิชก์ ผู้ช่วยศาสตราจารย์ ดร.มนชยา เจียงประดิษฐ์ และอาจารย์ ดร.สุภาวดี วิชิตชาญ ซึ่งเป็นผู้ทรงคุณวุฒิที่ให้ความอนุเคราะห์ ตรวจสอบงานวิจัย ถ่ายทอดความรู้ คำแนะนำ พร้อมทั้งข้อคิดที่เป็นประโยชน์อย่างยิ่ง และบุคคลท่านอื่น ๆ ที่ไม่ได้กล่าวนามทุกท่านที่ให้ข้อมูลต่าง ๆ คำแนะนำที่เอื้อต่อการทำงานวิจัยจนทำให้งานวิจัยนี้สำเร็จลุล่วงไปด้วยดี

ท้ายที่สุดขอขอบพระคุณภาควิชาคณิตศาสตร์ คณะวิทยาศาสตร์ มหาวิทยาลัยมหาสารคาม ที่เอื้อเฟื้อสถานที่ในการทำวิจัย และข้าพเจ้าใคร่ขอขอบพระคุณผู้ที่มีส่วนเกี่ยวข้องกับทุกท่านที่มีส่วนร่วมในการให้คำปรึกษา ทำให้วิจัยฉบับนี้จนเสร็จสมบูรณ์ ซึ่งข้าพเจ้าขอขอบพระคุณเป็นอย่างสูงไว้ ณ ที่นี้ด้วย

อธิปกรณ์ นาถมทอง

สารบัญ

หน้า

สารบัญตาราง

หน้า

บทที่ 1

บทนำ

1.1 หลักการและเหตุผล

ในการศึกษาคุณลักษณะของประชากรจะถูกเรียกว่า พารามิเตอร์ (Parameter) ซึ่งมีข้อจำกัดในการได้มาซึ่งข้อมูลไม่ว่าจะเป็นกำลังคน เวลา หรืองบประมาณ ส่งผลให้ไม่สามารถศึกษาข้อมูลทั้งหมดประชากรได้ จึงต้องอาศัยการสุ่มตัวอย่างหรือวิธีการเลือกตัวอย่าง (Sampling Technique) ขึ้นมาโดยกำหนดวิธีการเลือกตัวอย่างที่เหมาะสมกับลักษณะของประชากรที่ทำการศึกษาและเก็บข้อมูลมาเพียงบางหน่วยของประชากรซึ่งจะเรียกว่าเป็นการสำรวจ (Survey) จากนั้นนำข้อมูลที่เก็บได้จากตัวอย่างมาศึกษา ซึ่งคุณลักษณะของตัวอย่างอย่างจะถูกเรียกว่า ตัวสถิติ (Statistic) การนำข้อมูลที่ได้จากตัวอย่างนั้นไปอธิบายคุณลักษณะของประชากรนั้นต้องอาศัยสถิติอนุมาน (Inferential Statistics) ซึ่งอาจสรุปได้ในลักษณะของการประมาณค่า (Estimation) หรือการทดสอบสมมติฐาน (Hypothesis Testing) ในการประมาณค่าพารามิเตอร์ตัวหนึ่งอาจใช้ตัวประมาณที่ต่างกันได้ ซึ่งเป็นสิ่งที่จะต้องพิจารณาว่าข้อมูลที่เลือกมาเป็นตัวอย่างนั้นเหมาะสมกับตัวประมาณแบบใดมากที่สุด ซึ่งผลจากการประมาณค่าทางสถิติจะมีความถูกต้องและน่าเชื่อถือเพียงใดนั้นย่อมขึ้นอยู่กับวิธีการเลือกตัวอย่าง การเลือกพิจารณาตัวประมาณที่เหมาะสมมีหลักเกณฑ์ในการเลือก โดยสมบัติที่สำคัญ 2 ประการ ได้แก่ ความไม่เอนเอียง และ ความคลาดเคลื่อนต่ำ

การเลือกตัวอย่างเชิงสำรวจในการประมาณค่าประชากร จะทำการเลือกตัวอย่างจากประชากร มาวิเคราะห์และอธิบายถึงคุณลักษณะต่าง ๆ ของประชากร ด้วยวิธีการเลือกตัวอย่างแล้วนำข้อมูลไปประมาณค่าด้วยตัวประมาณ (Estimator) เมื่อพิจารณาการเลือกตัวอย่างสุ่มแบบง่าย (Simple random sampling: SRS) เป็นการเลือกตัวอย่างขนาด n หน่วย จากประชากรขนาด N หน่วย ซึ่งแต่ละหน่วยในประชากรมีโอกาสถูกเลือกเท่า ๆ กันสามารถทำได้ 2 วิธี คือ การเลือกตัวอย่างแบบง่ายไม่คืนที่ (Simple random sampling without replacement: SRSWOR) และการเลือกตัวอย่างแบบง่ายคืนที่ (Simple random sampling with replacement : SRSWR) ซึ่งในที่นี้ผู้วิจัยจะกล่าวถึงเฉพาะการเลือก ตัวอย่างเลือกตัวอย่างแบบง่ายไม่คืนที่เท่านั้น เนื่องจากเป็นวิธีการเลือกตัวอย่างที่มีประสิทธิภาพสูงกว่าวิธีการเลือกตัวอย่างแบบง่ายคืนที่

กำหนดให้ $Y_1, Y_2, \ldots, Y_N$ เป็นค่าของตัวแปรที่สนใจจะศึกษาในประชากรหน่วยที่ $i = 1, 2, 3, \ldots, N$ ดังนั้น ค่าเฉลี่ยของประชากร คือ $\bar{Y} = \dfrac{\sum Y_i}{N}$ และกำหนดให้ $y_1, y_2, \ldots, y_n$ เป็นค่าตัวแปรที่สนใจศึกษาในกลุ่มตัวอย่างหน่วยที่ $i = 1, 2, 3, \ldots, n$ ค่าเฉลี่ยของกลุ่มตัวอย่างใน

การเลือกตัวอย่างสุ่มแบบง่ายโดยไม่คืนที่ คือ $\bar{y} = \dfrac{\sum y_i}{n}$ ซึ่งเป็นตัวประมาณไม่เอนเอียง และเนื่องจาก $\bar{y}$ เป็นตัวประมาณที่ไม่เอนเอียงทำให้ค่าความเอนเอียง $(Bias)$ มีค่าเท่ากับศูนย์

ในการเลือกตัวอย่างเพื่อนำมาประมาณค่าของประชากรที่กล่าวมาข้างต้นนั้น เป็นการเลือกตัวอย่างสุ่มแบบง่ายโดยทำการเลือกตัวอย่างครั้งเดียวแบบไม่คืนที่ จะเห็นว่าเป็นการศึกษาตัวแปรที่สนใจ Y เพียงตัวแปรเดียว นอกจากตัวประมาณของประชากรดังกล่าวแล้วยังมีตัวประมาณอีกรูปแบบหนึ่งซึ่งนำเอาสารสนเทศของตัวแปรช่วย X มาช่วยเพิ่มประสิทธิภาพในการประมาณค่าเฉลี่ยของประชากรเพื่อให้ได้ค่าประมาณที่มีความแม่นยำมากขึ้นให้ได้ค่าที่ใกล้เคียงกับค่าพารามิเตอร์มากที่สุด และตัวแปรช่วย (Auxiliary variable) X ต้องเป็นตัวแปรที่ทราบค่าผลรวมทั้งหมดหรือค่าเฉลี่ยของประชากร โดยตัวแปรช่วย X นั้นอาจมีได้มากกว่า 1 ตัว ถ้าตัวแปร X มีความสัมพันธ์ในทิศทางเดียวกันกับตัวแปรที่ต้องการศึกษา Y ในระดับสูงจะเรียกตัวประมาณ ดังกล่าวว่า ตัวประมาณอัตราส่วน (Ratio Estimator) ซึ่งมีรูปแบบดังนี้

$$\bar{y}_R = \bar{y}\left(\frac{\bar{X}}{\bar{x}}\right) \quad \text{เมื่อ ทราบค่า } \bar{X}$$

โดย $\bar{X}$ คือ ค่าเฉลี่ยประชากรของตัวแปรช่วย

ในทางกลับกันถ้าตัวแปรช่วย X มีความสัมพันธ์ในทิศทางตรงกันข้ามกับตัวแปรที่ต้องการศึกษา Y ในระดับสูงจะเรียกตัวประมาณนี้ว่า ตัวประมาณผลคูณ (Product Estimator) มีรูปแบบดังนี้

$$\bar{y}_P = \bar{y}\left(\frac{\bar{x}}{\bar{X}}\right) \quad \text{เมื่อ ทราบค่า } \bar{X}$$

โดย $\bar{X}$ คือ ค่าเฉลี่ยประชากรของตัวแปรช่วย

อย่างไรก็ตาม ตัวประมาณอัตราส่วน (Ratio Estimator) หรือแม้ตัวประมาณผลคูณ (Product Estimator) นิยมใช้ในการประมาณค่าพารามิเตอร์ แต่ยังคงเป็นตัวประมาณที่เอนเอียง (Bias Estimator) ในการวิจัยครั้งนี้ผู้วิจัยจึงสนใจศึกษาทฤษฎีเกี่ยวกับตัวประมาณอัตราส่วนไม่เอนเอียงและตัวประมาณผลคูณไม่เอนเอียงดังนี้ ตัวประมาณอัตราส่วนที่พัฒนาโดย Tin (1965) ตัวประมาณผลคูณของ Robson (1957) และ ตัวประมาณไม่เอนเอียงของ Mahanty และ Mishra (2020)

โดย Tin (1965) เสนอตัวประมาณอัตราส่วนไม่เอนเอียง มีรูปแบบดังนี้

$$\bar{y}_{Tin} = \bar{y}\frac{\bar{X}}{\bar{x}}\left[1 + f\left(\frac{s_{yx}}{\bar{y}\,\bar{x}} - \frac{s_x^2}{\bar{x}^2}\right)\right]$$

Robson (1957) เสนอตัวประมาณผลคูณไม่เอนเอียง มีรูปแบบดังนี้

$$\bar{y}_{Rob} = \frac{\bar{y}\,\bar{x}}{\bar{X}} - f\frac{s_{yx}}{\bar{X}}$$

Mahanty & Mishra (2020) เสนอตัวประมาณไม่เอนเอียง มีรูปแบบดังนี้

$$\bar{y}_{MU} = \lambda_0\left(\bar{y}\frac{\bar{x}}{\bar{X}} - f\frac{s_{xy}}{\bar{X}}\right) + \lambda_1\bar{x}$$

เมื่อ $f=\frac{1}{n}-\frac{1}{N}$ และ

λ_0 และ λ_1 เป็นค่าคงที่ที่เหมาะสม 2 ค่าที่ทำให้ตัวประมาณ $\bar{y}_{MU}$ ไม่เอนเอียง

ในการประมาณค่าเฉลี่ยของตัวแปรที่สนใจศึกษา Y โดยใช้ตัวประมาณอัตราส่วน ตัวประมาณผลคูณ นั้นต้องทราบค่าเฉลี่ยประชากรของตัวแปรช่วย $\bar{X}$ ในบางครั้งไม่สามารถทราบค่า $\bar{X}$ ได้ แต่สามารถเก็บค่าข้อมูลของตัวแปรช่วย X และตัวแปรที่สนใจศึกษา Y ได้เมื่อทำการเลือกตัวอย่างมาแล้ว

ในบางสถานการณ์ตัวแปรที่สนใจศึกษา Y ยากต่อการเก็บข้อมูลหรือค่าใช้จ่ายในการเก็บข้อมูลมีราคาแพงมาก แต่ถ้ามีตัวแปรอื่นที่มีความสัมพันธ์กับตัวแปรที่สนใจศึกษาและสามารถเก็บข้อมูลได้ง่ายกว่าหรือมีค่าใช้จ่ายที่ถูกกว่า เรียกตัวแปรนี้ว่าตัวแปรช่วย X โดยทำการเก็บ ข้อมูลตัวแปรช่วย X จากตัวอย่างขนาดใหญ่กว่าข้อมูลจากตัวแปรที่สนใจศึกษา Y เรียกวิธีการ เลือกตัวอย่างนี้ว่า "การเลือกตัวอย่างสองเฟส" (Double sampling หรือ Two-phase sampling) วิธีการเลือกตัวอย่างสองเฟสทำได้ดังนี้

เฟสที่ 1 ประชากรขนาด N หน่วย เลือกตัวอย่างขนาด $n^{(1)}$ หน่วย จากนั้นเก็บข้อมูลเฉพาะตัวแปรช่วย X เท่านั้น

เฟสที่ 2 จากตัวอย่างขนาด $n^{(1)}$ หน่วยในเฟสที่ 1 เลือกตัวอย่างขนาด $n^{(2)}$ หน่วยจากนั้นเก็บข้อมูลตัวแปรที่สนใจศึกษา Y และตัวแปรช่วย X

โดยในการวิจัยครั้งนี้ผู้วิจัยได้ศึกษาตัวประมาณแบบอัตราส่วน และตัวประมาณผลคูณ ในการเลือกตัวอย่างสองเฟส ซึ่งมีรูปแบบดังนี้

ตัวประมาณอัตราส่วนในการเลือกตัวอย่างสองเฟส

$$\bar{y}_R^{(2)}=\bar{y}^{(2)}\left(\frac{\bar{x}^{(1)}}{\bar{x}^{(2)}}\right)$$

และตัวประมาณผลคูณในการเลือกตัวอย่างสองเฟส

$$\bar{y}_P^{(2)}=\bar{y}^{(2)}\left(\frac{\bar{x}^{(2)}}{\bar{x}^{(1)}}\right)$$

เมื่อ $\bar{x}^{(1)}=\frac{1}{n^{(1)}}\sum_{i=1}^{n^{(1)}}x_i$ คือ ค่าประมาณค่าเฉลี่ยของตัวแปรช่วย X ในเฟสที่ 1

y_i คือ ค่าสังเกตของตัวแปรที่สนใจศึกษาหน่วยที่ i

x_i คือ ค่าสังเกตของตัวแปรช่วย หน่วยที่ i

$n^{(1)}$ คือ ขนาดตัวอย่างของการเลือกตัวอย่างเฟสที่ 1

$n^{(2)}$ คือ ขนาดตัวอย่างของการเลือกตัวอย่างเฟสที่ 2

ดังนั้น ในการวิจัยครั้งนี้ผู้วิจัยสนใจศึกษาทฤษฎีเกี่ยวกับตัวประมาณอัตราส่วนไม่เอนเอียง และตัวประมาณผลคูณไม่เอนเอียง ที่เสนอโดย Tin (1965) Robson (1957) และ Mahanty & Mishra (2020) ภายใต้การเลือกตัวอย่างสองเฟส และเปรียบเทียบประสิทธิภาพตัวประมาณที่ศึกษากับตัวประมาณอัตราส่วนและตัวประมาณผลคูณแบบทั่วไป

1.2 วัตถุประสงค์ของการวิจัย

1.2.1 เพื่อศึกษาและพัฒนาตัวประมาณไม่เอนเอียงโดยการใช้สารสนเทศของตัวแปรช่วยสำหรับค่าเฉลี่ยประชากรจำกัด ภายใต้การเลือกตัวอย่างสองเฟส

1.2.2 เพื่อเปรียบเทียบประสิทธิภาพของตัวประมาณที่ศึกษากับตัวประมาณอัตราส่วนและตัวประมาณผลคูณแบบทั่วไป

1.3 ขอบเขตของการวิจัย

1.3.1 การวิจัยครั้งนี้ผู้วิจัยศึกษาตัวประมาณภายใต้วิธีการเลือกตัวอย่างสองเฟส

1.3.2 ในการเปรียบเทียบประสิทธิภาพของตัวประมาณ ผู้วิจัยเปรียบเทียบประสิทธิภาพโดยใช้ค่าความแปรปรวน (Variance) เป็นเกณฑ์

1.3.3 ผู้วิจัยวิเคราะห์ข้อมูลผ่านการจำลองข้อมูลโดยใช้โปรแกรม RStudio ในการจำลองข้อมูลสร้างกลุ่มประชากรขึ้นมา โดยสร้างประชากรให้มีการแจกแจงปรกติและกำหนดความสัมพันธ์กันในระดับต่าง ๆ

สร้างประชากรขนาด $N=1,000$ หน่วย โดย $Y \sim N\left(100, 20^2\right)$ และ $X \sim N\left(500,100^2\right)$

กรณี Y และ X มีความสัมพันธ์เชิงเส้นตรงในทิศทางเดียวกัน กำหนดค่าความสัมพันธ์ของประชากร 3 ระดับ คือ

1 ประชากรมีความสัมพันธ์กันสูง $\rho_{YX}=0.8$

2 ประชากรมีความสัมพันธ์กันปานกลาง $\rho_{YX}=0.6$

3 ประชากรมีความสัมพันธ์กันน้อย $\rho_{YX}=0.3$

กรณี Y และ X มีความสัมพันธ์เชิงเส้นตรงในทิศทางตรงกันข้าม กำหนดค่าความสัมพันธ์ของประชากร 3 ระดับ คือ

1 ประชากรมีความสัมพันธ์กันสูง $\rho_{YX}=-0.8$

2 ประชากรมีความสัมพันธ์กันปานกลาง $\rho_{YX}=-0.6$

3 ประชากรมีความสัมพันธ์กันน้อย $\rho_{YX}=-0.3$

1.4 ประโยชน์ที่คาดว่าจะได้รับ

1.4.1 เป็นแนวทางในการเลือกใช้ตัวประมาณไม่เอนเอียงเมื่อใช้สารสนเทศของตัวแปรช่วยเพื่อประมาณค่าเฉลี่ยของประชากรภายใต้การเลือกตัวอย่างสองเฟส

1.4.1 เพื่อได้ตัวประมาณที่มีประสิทธิภาพในการประมาณค่าเฉลี่ยของประชากรในการเลือกตัวอย่างสองเฟส

1.5 นิยามศัพท์เฉพาะ

1.5.1 ประชากร (Population) คือ หน่วยทุกหน่วยของสิ่งที่เราสนใจจะทำการศึกษา เป็นหน่วยที่สามารถให้ข้อมูลต่าง ๆ แก่เราได้

1.5.2 ตัวอย่าง (Sample) คือ หน่วยตัวอย่างส่วนหนึ่งของประชากรที่ถูกเลือกขึ้นมาเพื่อใช้เป็นตัวแทนในการศึกษาหารายละเอียดข้อเท็จจริงที่ต้องการเกี่ยวกับประชากรนั้น เช่น เกษตรกรกลุ่มหนึ่งที่ถูกเลือกขึ้นมาเป็นตัวแทนของเกษตรกรทั้งหมดของประเทศ จากตัวอย่างที่เลือกมาได้จะเก็บรวบรวมข้อมูลที่ต้องการและนำไปคำนวณเพื่อใช้ในการประมาณค่าพารามิเตอร์

1.5.3 ตัวประมาณไม่เอนเอียง (Unbiased Estimator) คือค่าคาดหวังของตัวสถิติ ต้องเท่ากับ ค่าพารามิเตอร์

1.5.4 สารสนเทศของตัวแปรช่วย (Auxiliary Information) คือ ตัวแปรที่นำมาช่วยเสริมค่าของตัวแปรที่สนใจศึกษาในการประมาณค่าพารามิเตอร์เพื่อให้มีประสิทธิภาพมากขึ้น

1.5.5 การเลือกตัวอย่างสองเฟส (Two-Phase Sampling) คือ วิธีการเลือกตัวอย่างที่ดำเนินการเลือกเป็น 2 ตอน กล่าวคือ ตอนที่ 1 เลือกตัวอย่างขนาดใหญ่ (มีจำนวนหน่วยตัวอย่างมาก) มาโดยวิธีใดวิธีหนึ่ง แล้วเก็บรวบรวมข้อมูลเบื้องต้น เพื่อใช้ประโยชน์ในการเลือกตัวอย่างในตอนที่ 2

บทที่ 2
เอกสารและงานวิจัยที่เกี่ยวข้อง

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อศึกษาตัวประมาณไม่เอนเอียงโดยการใช้สารสนเทศของตัวแปรช่วยสำหรับค่าเฉลี่ยประชากรจำกัด ภายใต้การเลือกตัวอย่างสองเฟส และเปรียบเทียบประสิทธิภาพของตัวประมาณที่ศึกษากับตัวประมาณอัตราส่วนและตัวประมาณผลคูณแบบทั่วไป งานวิจัยจึงมีความจำเป็นที่จะต้องทำการทบทวบวรรณกรรมที่เกี่ยวข้องต่าง ๆ โดยแบ่งออกเป็น

2.1 ทฤษฎีที่เกี่ยวข้อง
- 2.1.1 ประชากร (Population)
- 2.1.2 ตัวอย่าง (Sample)
- 2.1.3 การประมาณค่า (Estimation)
- 2.1.4 การเลือกตัวอย่างสุ่มแบบง่ายไม่คืนที่ (Simple random sampling without replacement) และตัวประมาณค่าเฉลี่ยของประชากร
- 2.1.5 การเลือกตัวอย่างสองเฟส (Two-Phase Sampling) และตัวประมาณค่าเฉลี่ยของประชากร
- 2.1.6 ตัวประมาณไม่เอนเอียง (Unbiased Estimator)
- 2.1.7 การเปรียบเทียบประสิทธิภาพของตัวประมาณ (Comparison of Efficiency)

2.2 งานวิจัยที่เกี่ยวข้อง
- 2.2.1 งานวิจัยภายในประเทศ
- 2.2.2 งานวิจัยต่างประเทศ

2.1 ทฤษฎีที่เกี่ยวข้อง

2.1.1 ประชากร (Population)

ประชากร หมายถึง หน่วยทุกหน่วยของสิ่งที่เราสนใจจะทำการศึกษา เป็นหน่วยที่สามารถให้ข้อมูลต่าง ๆ แก่เราได้ เช่น สนใจรายได้เฉลี่ยของบุคลากรในมหาวิทยาลัยมหาสารคาม ดังนั้นประชากรคือ บุคลากรทุกคนที่ทำงานอยู่ในมหาวิทยาลัยมหาลัยมหาสารคาม ประชากรแบ่งเป็น 2 ชนิด คือ

ประชากรจำกัด (Finite population) หมายถึง ประชากรที่ประกอบด้วยหน่วยตัวอย่างแน่นอน

ประชากรไม่จำกัด (Infinite population) หมายถึง ประชากรที่ประกอบด้วยหน่วยตัวจำนวนไม่แน่นอนหรือมีจำนวนอนันต์นั่นเอง

ขนาดของประชากร (Population size) เป็นพารามิเตอร์ตัวหนึ่งของประชากร ขนาดของประชากรจะแสดงถึงจำนวนหน่วยตัวอย่างทั้งหมดของประชากรหนึ่ง เช่น ในการศึกษาหนูนาในท้องที่

แห่งหนึ่งอาจมีความต้องการทราบจำนวนของหนูนาทั้งหมดในท้องที่แห่งนั้นว่ามีจำนวนเท่าไร เพื่อใช้ข้อมูลเป็นเบื้องต้นในการวางแผนกำจัดต่อไปหรือในการหาจำนวนของสัตว์สงวนพันธุ์ในป่าแห่งหนึ่ง เจ้าหน้าที่ป่าสงวนอาจจะต้องการทราบจำนวนของสัตว์ป่าสงวนพันธุ์ทั้งหมดที่ยังเหลืออยู่ว่ามีจำนวนเท่าไร เพื่อจะใช้เป็นข้อมูลเบื้องต้นสำหรับการวางแผนการอนุรักษ์พันธ์สัตว์ป่าชนิดนั้นต่อไป

พารามิเตอร์ (Parameter) หมายถึง ค่าคงตัวที่แสดงถึงลักษณะเฉพาะบางประการของประชากรหรือเป็นฟังก์ชันที่เกี่ยวข้องกับข้อมูลจากทุกหน่วยตัวอย่างของประชากร พารามิเตอร์นี้มักไม่ทราบค่า ดังนั้นในการศึกษาจึงต้องทำการประมาณค่าพารามิเตอร์ว่ามีค่าเท่าใด

คุณลักษณะบางประการของประชากร เราจะเรียกว่า ค่าพารามิเตอร์ (Parameter) ซึ่งมีดังต่อไปนี้

ให้ N คือ ขนาดของประชากร หมายถึง จำนวนหน่วยตัวอย่างทั้งหมดของประชากร

Y_i คือ ค่าคงที่แสดงคุณลักษณะที่ i ของประชากร $i = 1, 2, 3, \ldots, N$

1) ยอดรวมของประชากร (Population total)

ใช้สัญลักษณ์ Y คือ ยอดรวมของประชากร โดย $Y = \sum_{i=1}^{N} Y_i$

2) ค่าเฉลี่ยของประชากร (Population mean)

ใช้สัญลักษณ์ $\bar{Y}$ คือ ค่าเฉลี่ยของประชากร โดย $\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i$

3) ความแปรปรวนของประชากร (Population variance)

ใช้สัญลักษณ์ S_Y^2 คือ ความแปรปรวนของประชากร โดย $S_Y^2 = \frac{1}{N}\sum_{i=1}^{N}\left(Y_i - \bar{Y}\right)^2$

S_Y คือ ส่วนเบี่ยงเบนมาตรฐานของประชากร โดย $S_Y = \sqrt{S_Y^2}$

2.1.2 ตัวอย่าง (Sample)

ตัวอย่าง (Sample) คือ หน่วยตัวอย่างส่วนหนึ่งของประชากรที่ถูกเลือกขึ้นมาเพื่อใช้เป็นตัวแทนในการศึกษาหารายละเอียดข้อเท็จจริงที่ต้องการเกี่ยวกับประชากรนั้น เช่น เกษตรกรกลุ่มหนึ่งที่ถูกเลือกขึ้นมาเป็นตัวแทนของเกษตรกรทั้งหมดของประเทศ ต้นไม้จำนวนหนึ่งที่ถูกเลือกขึ้นมาเป็นตัวแทนของต้นไม้ทั้งหมดในป่าแห่งหนึ่ง ร้านค้าส่วนหนึ่งที่ถูกเลือกขึ้นมาเป็นตัวแทนของร้านค้าทั้งหมด เป็นต้น จากตัวอย่างที่เลือกมาได้จะเก็บรวบรวมข้อมูลที่ต้องการและนำไปคำนวณเพื่อใช้ในการประมาณค่าพารามิเตอร์

ตัวสถิติ (Statistic) คือฟังก์ชันของค่าสังเกตที่วัดมาจากหน่วยตัวอย่างต่าง ๆ ที่ถูกเลือกขึ้นมาเป็นตัวอย่าง ฟังก์ชันดังกล่าวจะไม่มีตัวพารามิเตอร์อื่นใดที่ยังไม่ทราบค่าติดอยู่เลย ซึ่งอาจกล่าวได้ว่าตัวสถิติก็คือสูตรสำหรับการคำนวณตัวสถิติที่ได้จะใช้เป็นตัวประมาณ (Estimator) ของพารามิเตอร์ ส่วนค่าสถิติ คือ ค่าของตัวสถิติที่คำนวณออกมาเป็นตัวเลขแล้วค่าสถิติจะใช้เป็นค่าประมาณ (Estimate) ของพารามิเตอร์ที่ยังไม่ทราบค่าต่อไป ตัวสถิติที่สำคัญมีดังต่อไปนี้

คุณลักษณะบางประการของหน่วยตัวอย่าง เราจะเรียกว่า ตัวสถิติ (Statistic) ที่สำคัญจะมีดังต่อไปนี้

ให้ n คือ จำนวนของหน่วยตัวอย่างที่ถูกเลือกมาเป็นตัวแทนของประชากร

y_i คือ ค่าสังเกตที่ i ของหน่วยตัวอย่าง $i = 1, 2, 3, \ldots, n$

1) ยอดรวมของตัวอย่าง (Sample total)

ใช้สัญลักษณ์ y คือ ยอดรวมของตัวอย่าง โดย $y = \sum_{i=1}^{n} y_i$

2) ค่าเฉลี่ยของตัวอย่าง (Sample mean)

ใช้สัญลักษณ์ $\bar{y}$ คือ ค่าเฉลี่ยของตัวอย่าง โดย $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$

3) ความแปรปรวนของตัวอย่าง (Sample variance)

ใช้สัญลักษณ์ s_y^2 คือ ความแปรปรวนของตัวอย่าง โดย $s_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2$

s_y คือ ส่วนเบี่ยงเบนมาตรฐานของตัวอย่าง โดย $s_y = \sqrt{s_y^2}$

ขนาดตัวอย่าง (Sample size) หมายถึง จำนวนหน่วยในประชากรที่ถูกเลือกขึ้นมาให้เป็นตัวแทนของประชากรที่ต้องการศึกษา โดยหน่วยที่ถูกเลือกขึ้นมานั้นจะเป็นตัวแทนเรียกว่า หน่วยตัวอย่าง (sampling unit) สัญลักษณ์ใช้แทนขนาดตัวอย่าง คือ n ค่า n ได้จากการคำนวณและอยู่ในวิสัยสามารถปฏิบัติงานได้ โดยค่า n จะมีค่าน้อยกว่าค่า N (ภาควิชาคณิตศาสตร์, 2558)

2.1.3 การประมาณค่า (Estimation)

การประมาณค่าเป็นการหาว่าพารามิเตอร์ในประชากรมีค่าเป็นเท่าใด ซึ่งถือว่าเป็นการประมาณค่าพารามิเตอร์ โดยใช้สถิติที่มีคุณสมบัติของการเป็นตัวประมาณที่ดี คือ ความไม่เอนเอียง (Unbiasedness) มีความแปรปรวนต่ำที่สุด (Minimum Variance) ความพอเพียง (Sufficiency) ความมีประสิทธิภาพ (Efficiency) และความคงเส้นคงวา (Consistency) เป็นต้น ซึ่งการประมาณค่าแบ่งออกเป็น 2 แบบ ดังนี้

2.1.3.1 การประมาณค่าแบบจุด (Point Estimation)

เป็นการประมาณค่าพารามิเตอร์ที่ได้เพียงค่าเดียวเท่านั้น โดยไม่มีหลักประกันว่าค่าประมาณที่ได้นั้นมีค่าใกล้เคียงกับค่าจริงมากน้อยเพียงใด ซึ่งการใช้สถิติเป็นฟังก์ชันของตัวอย่างสุ่ม $t\left(X_1, \ldots, X_n\right)$ โดยการสังเกตค่าของตัวแปร $X_1, \ldots, X_n$ สมมติว่าได้ $x_1, \ldots, x_n$ แล้วใช้ค่าของ $\hat{\theta} = t\left(x_1, \ldots, x_n\right)$ เป็นค่าประมาณของ θ ดังนั้นการประมาณค่าแบบนี้จึงมีความคลาดเคลื่อนได้มาก

2.1.3.2 การประมาณค่าแบบช่วง (Interval Estimation)

เป็นที่ทราบกันดีว่าตัวอย่างสุ่มที่นำมาศึกษาจะให้ค่าที่เรียกว่า ค่าสถิติ และค่าที่ศึกษาจากทุกหน่วยในประชากรจะเรียกว่า ค่าพารามิเตอร์ ดังนั้นโอกาสที่ค่าสถิติจะมีค่าเท่ากับค่าพารามิเตอร์หรือคลาดเคลื่อนจากค่าพารามิเตอร์ก็สามารถเกิดขึ้นได้ ซึ่งความคลาดเคลื่อนที่เกิดขึ้นจากความผันแปรที่เกิดจากการเลือกตัวอย่างหรือการสุ่มตัวอย่าง จึงต้องทำให้มีการประมาณค่าแบบช่วงควบคู่ไปกับการประมาณค่าแบบจุดขึ้น

การประมาณค่าแบบช่วง เป็นการประมาณค่าพารามิเตอร์ที่ระบุค่าในรูปของช่วง (L,U) ซึ่งเชื่อมั่นว่าครอบคลุมค่าพารามิเตอร์ที่เราต้องการประมาณค่า แม้จะไม่สามารถรับประกันได้ว่าช่วงประมาณ (L,U) นี้ครอบคลุมค่าที่แท้จริงของพารามิเตอร์หรือไม่ แต่การประมาณค่าแบบช่วงจะครอบคลุมค่าพารามิเตอร์ได้มากกว่าแบบจุด ทำให้มีโอกาสที่จะคลาดเคลื่อนไปจากค่าที่แท้จริงน้อยกว่าการประมาณค่าแบบจุด

ช่วงความเชื่อมั่น (Confidence interval) โดยที่

L คือ ขีดจำกัดล่างของช่วงความเชื่อมั่น (Lower confidence limit)

U คือ ขีดจำกัดบนของช่วงความเชื่อมั่น (Upper confidence limit)

และมี ระดับความเชื่อมั่น (Level of confidence: $1-\alpha$) คือ ความน่าจะเป็นที่ช่วงประมาณนี้จะครอบคลุมค่าพารามิเตอร์

ถ้า θ เป็นพารามิเตอร์ที่เราต้องการประมาณค่า จะได้ว่า ช่วงความเชื่อมั่น $(1-\alpha)100\%$ ของ θ คือ ช่วง (L , U) หรือ L < θ < U ซึ่งมีระดับความเชื่อมั่น $(1-\alpha)$ หรือสามารถเขียนในรูปของความน่าจะเป็นได้ดังนี้คือ P(L < θ < U) = $1-\alpha$

ระดับความเชื่อมั่นที่นิยมเลือกใช้มากที่สุดคือ 0.90 0.95 และ 0.99 เช่น ถ้ากำหนด $\alpha = 0.05$ คือ ต้องการสร้างช่วงความเชื่อมั่น $(1-\alpha)100\% = 95\%$ ทำการเลือกตัวอย่างขนาด n จากประชากรชุดเดิม 100 ครั้ง เพื่อสร้างช่วงความเชื่อมั่น จะมีช่วงความเชื่อมั่น 95 ช่วงความเชื่อมั่น ที่ครอบคลุมค่าพารามิเตอร์ที่แท้จริง และอีก 5 ช่วงความเชื่อมั่นที่ไม่ครอบคลุมค่าพารามิเตอร์

ถ้าช่วงความเชื่อมั่นกว้าง หมายถึงโอกาสที่ค่าประมาณจะคลุมค่าพารามิเตอร์ก็มากตามไปด้วยแต่ไม่ได้หมายความว่าจะเป็นสิ่งที่ดี เนื่องจากเราไม่ทราบค่าที่แท้จริงของค่าพารามิเตอร์ ถ้าช่วงกว้างมาก ๆ ประมาณค่าอย่างไรก็ยังคลุมค่าพารามิเตอร์อยู่นั่นเอง นั่นคือ ถ้าช่วงกว้างจะทำให้ความเที่ยงตรงของการประมาณค่าลดลง ในทางกลับกัน ถ้าช่วงแคบก็จะส่งผลต่อการเพิ่มความเที่ยงตรงของการประมาณค่า ดังนั้นจุดที่ดีที่สุดคือ การประมาณค่าที่ L และ U ให้ช่วง (L , U) สั้นที่สุดซึ่งสามารถทำได้ 2 วิธีคือ ลดระดับความเชื่อมั่น หรือ เพิ่มขนาดตัวอย่าง

2.1.4 การเลือกตัวอย่างสุ่มแบบง่ายไม่คืนที่ (Simple random sampling without replacement) และตัวประมาณค่าเฉลี่ยของประชากร

การเลือกตัวอย่างสุ่มแบบง่าย (Simple random sampling) เป็นวิธีการเลือกหน่วยตัวอย่าง จากประชากรทั้งหมดขนาด N หน่วย มาจำนวน n หน่วย โดยกำหนดให้ตัวอย่างแต่ละตัวอย่างที่จะเป็นไปได้มีโอกาสจะถูกเลือกเท่า ๆ กัน

วิธีการการเลือกตัวอย่างสุ่มแบบง่าย ทำได้โดยการเลือกหน่วยตัวอย่างมาทีละหน่วยโดยสุ่ม โดยให้การเลือกแต่ละครั้งมีความน่าจะเป็นเท่า ๆ กัน ดำเนินการได้ 2 แบบคือ

การเลือกตัวอย่างแบบไม่คืนที่ (Simple random sampling without replacement: SRSWOR) หน่วยที่ถูกสุ่มมาแล้วจะไม่คืนกลับไปในประชากร และจะไม่ถูกเลือกได้อีกจะได้ว่า

จำนวนตัวอย่างที่เป็นไปได้ทั้งหมดเท่ากับ ${}^{N}C_n$ ดังนั้นความน่าจะเป็นที่ตัวอย่างแต่ละตัวอย่างจะถูกเลือกเท่ากับ $\frac{1}{{}^{N}C_n}$

การเลือกตัวอย่างแบบคืนที่ (Simple random sampling with replacement: SRSWR) หน่วยที่ถูกเลือกแล้วจะสามารถถูกเลือกได้อีก จะได้ว่าจำนวนตัวอย่างที่เป็นไปได้ทั้งหมดเท่ากับ N^n ดังนั้นความน่าจะเป็นที่ตัวอย่างแต่ละตัวอย่างจะถูกเลือกเท่ากับ $\frac{1}{N^n}$ (นิภาพร ชุติมันต์ (2557))

ตัวประมาณค่าเฉลี่ยของประชากรภายใต้การเลือกตัวอย่างแบบง่ายไม่คืนที่

ตัวประมาณอัตราส่วน (Ratio Estimator)

ถ้าหากว่าตัวแปร X มีความสัมพันธ์ในทิศทางเดียวกันกับตัวแปรที่ต้องการศึกษา Y ในระดับสูงและเรียกตัวประมาณดังกล่าวว่า ตัวประมาณอัตราส่วน (Ratio Estimator)

ตัวประมาณอัตราส่วนแบบทั่วไป (Conventional Ratio Estimator)

เมื่อทำการเลือกตัวอย่างขนาด n จากประชากรขนาด N ด้วยวิธีการเลือกตัวอย่างแบบง่ายไม่คืนที่ ประมาณ $\bar{Y}$ ด้วย $\bar{y}$ อย่างไรก็ตาม ถ้าทราบสารสนเทศของตัวแปรช่วย X เช่น ทราบค่าเฉลี่ย $\bar{X}$ ก็สามารถ ใช้สารสนเทศเหล่านี้มาช่วยในการประมาณค่า $\bar{Y}$ ให้มีความแม่นยำสูงขึ้น โดยที่ตัวแปรช่วยต้องเป็นตัวแปรที่มีความสัมพันธ์เชิงเส้นตรงทางบวกกับตัวแปรที่สนใจศึกษาในระดับสูง (Irfan et al., 2020) กำหนดให้

$$R = \frac{\sum_{i=1}^{N} Y_i}{\sum_{i=1}^{N} X_i} = \frac{\bar{Y}}{\bar{X}}$$

เรียก R คืออัตราส่วนของประชากร ซึ่งเป็นพารามิเตอร์ที่ต้องการประมาณค่า ซึ่งจะประมาณ R ด้วยตัวประมาณ $\hat{R}$ โดยที่

$$\hat{R} = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i} = \frac{\bar{y}}{\bar{x}}$$

สำหรับตัวประมาณแบบอัตราส่วนของค่าเฉลี่ยประชากรนั้น ตัวแปรช่วย (Auxiliary variables, x_i) ควรมีความสัมพันธ์เชิงเส้นตรงในทิศทางเดียวกันกับตัวแปรที่สนใจศึกษา (Interest variable, y_i) โดยตัวประมาณแบบอัตราส่วนเพื่อประมาณค่าเฉลี่ยประชากรคือ

$$\bar{y}_R = \hat{R}\bar{X} \quad \text{เมื่อ ทราบค่า } \bar{X}$$

โดยมีค่าความแปรปรวนดังนี้

$$V(\bar{y}_R) \approx \left(f\bar{Y}^2\right)\left[C_Y^2 + C_X^2 - 2\rho_{YX}C_YC_X\right]$$

เมื่อค่า $f = \frac{1}{n} - \frac{1}{N},\ C_Y^2 = \frac{S_y^2}{\bar{Y}^2},\ C_X^2 = \frac{S_x^2}{\bar{X}^2},\ \rho_{YX} = \frac{S_{yx}}{S_yS_x},\ S_y^2 = \frac{1}{(N-1)}\sum_{i=1}^{N}\left(y_i - \bar{Y}\right)^2,$

$$S_x^2 = \frac{1}{(N-1)}\sum_{i=1}^{N}\left(x_i - \bar{X}\right)^2,\ S_{yx}^2 = \frac{1}{(N-1)}\sum_{i=1}^{N}\left(y_i - \bar{Y}\right)\left(x_i - \bar{X}\right)$$

เมื่อ f คือ เศษส่วนการเลือกตัวอย่าง (Sampling fraction)

C_Y คือ สัมประสิทธิ์ความแปรผันของประชากรของตัวแปร Y

C_X คือ สัมประสิทธิ์ความแปรผันของประชากรของตัวแปร X

ρ_{YX} คือ สัมประสิทธิ์สหสัมพันธ์ของตัวแปร X และ ตัวแปร Y

ตัวประมาณอัตราส่วนที่เสนอโดย Tin (1965)

ได้นำเสนอตัวประมาณอัตราส่วนไม่เอียงในรูปแบบใหม่ ที่มีค่าความคลาดเคลื่อนต่ำ เพื่อใช้ในการประมาณค่าเฉลี่ยของประชากรจำกัด โดย

$$\bar{y}_{Tin} = \bar{y}\frac{\bar{X}}{\bar{x}}\left[1 + f\left(\frac{s_{yx}}{\bar{y}\,\bar{x}} - \frac{s_x^2}{\bar{x}^2}\right)\right]$$

โดยมีค่าความแปรปรวนดังนี้

$$V\left(\bar{y}_{Tin}\right) \approx f\bar{Y}^2\left[C_Y^2 + C_X^2 - 2\rho_{YX}C_YC_X\right]$$

ตัวประมาณแบบผลคูณ (Product Estimator)

ถ้าหากว่าตัวแปร X มีความสัมพันธ์ในทิศทางตรงกันข้ามกันกับตัวแปรที่ต้องการศึกษา Y ในระดับสูงและเรียกตัวประมาณดังกล่าวว่า ตัวประมาณผลคูณ (Product Estimator)

ตัวประมาณผลคูณแบบทั่วไป (Conventional Product Estimator)

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

โดยมีค่าความแปรปรวนดังนี้ $V\left(\bar{y}_P\right) \approx \left(f\bar{Y}^2\right)\left[C_Y^2 + C_X^2 + 2\rho_{yx}C_YC_X\right]$

ตัวประมาณผลคูณที่เสนอโดย Robson (1957)

ได้นำเสนอตัวประมาณผลคูณไม่เอียงในรูปแบบใหม่ ที่มีค่าความคลาดเคลื่อนต่ำ เพื่อใช้ในการประมาณค่าเฉลี่ยของประชากรจำกัด โดย

$$\bar{y}_{Rob} = \frac{\bar{y}\,\bar{x}}{\bar{X}} - f\frac{s_{yx}}{\bar{X}}$$

โดยมีค่าความแปรปรวนดังนี้

$$V\left(\bar{y}_{Rob}\right) \approx f\bar{Y}^2\left[C_Y^2 + C_X^2 + 2\rho_{YX}C_YC_X\right]$$

ตัวประมาณที่เสนอโดย Mahanty & Mishra (2020)

ได้สร้างตัวประมาณที่ไม่เอนเอียงขึ้นโดยใช้ผลรวมเชิงเส้นของตัวประมาณตัวแปรที่ศึกษาและค่าเฉลี่ยต่อหน่วยของตัวแปรช่วยภายใต้การสุ่มตัวอย่างอย่างง่ายแบบไม่คืนที่ เมื่อพิจารณาผลรวมเชิงเส้นของตัวประมาณที่ไม่เอนเอียงของร็อบสันและตัวประมาณค่าเฉลี่ยต่อหน่วยของตัวแปรช่วย โดย

$$\bar{y}_{MU} = \lambda_0\left(\bar{y}\frac{\bar{x}}{\bar{X}} - f\frac{s_{xy}}{\bar{X}}\right) + \lambda_1\bar{x}$$

โดย λ_0 และ λ_1 เป็นค่าคงที่ที่เหมาะสม 2 ค่าที่ทำให้ตัวประมาณ $\bar{y}_{MU}$ ไม่เอนเอียง ดังนั้น จึงตั้งเงื่อนไขที่ไม่เอนเอียงขึ้น โดย

$$E(\bar{y}_{MU}) = E[\lambda_0 \bar{y}_{Rob} + \lambda_1 \bar{x}] = \bar{Y}$$

$$\Rightarrow \lambda_0 E(\bar{y}_{Rob}) + \lambda_1 E(\bar{x}) = \bar{Y}$$

$$\Rightarrow \lambda_0 \bar{Y} + \lambda_1 \bar{X} = \bar{Y}$$

$$\Rightarrow (\lambda_0 - 1)\bar{Y} + \lambda_1 \bar{X} = 0$$

โดยได้ค่าความแปรปรวนของตัวประมาณที่ไม่เอนเอียงที่เสนอ โดยพิจารณาจากเงื่อนไข ค่าของ λ_0 และ λ_1 ซึ่งเป็นค่าที่ทำให้ตัวประมาณมีความไม่เอนเอียง

ความแปรปรวนของตัวประมาณ $\bar{y}_{MU}$

$$V(\bar{y}_{MU}) = \left[\lambda_0^2 V(\bar{y}_{Rob}) + \lambda_1^2 V(\bar{x}) + 2\lambda_0\lambda_1 Cov(\bar{x}, \bar{y}_{Rob})\right]$$

$$= \lambda' S_2 \lambda$$

เมื่อ $\lambda' = [\lambda_0 \quad \lambda_1]$, $S_2 = \begin{bmatrix} S_{00} & S_{01} \\ S_{10} & S_{11} \end{bmatrix}$ และ $\lambda = \begin{bmatrix} \lambda_0 \\ \lambda_1 \end{bmatrix}$

นอกจากนี้ $S_{00} = V(\bar{y}_{Rob})$, $S_{10} = S_{01} = Cov(\bar{x}, \bar{y}_{Rob})$ และ $S_{11} = V(\bar{x})$

$$S_{00} = V(\bar{y}_{Rob}) = f\bar{Y}^2\left[C_Y^2 + C_X^2 + 2\rho_{YX} C_Y C_X\right]$$

$$S_{10} = S_{01} = Cov(\bar{x}, \bar{y}_{Rob}) = f\overline{YX}\left[\rho_{YX} C_Y C_X + C_X^2\right]$$

$$S_{11} = V(\bar{x}) = f\bar{X}^2 C_X^2$$

เมื่อ $$\lambda_0 = \frac{R^2 V(\bar{x}) - RCov(\bar{x}, \bar{y}_{Rob})}{V(\bar{y}_{Rob}) + R^2 V(\bar{x}) - 2RCov(\bar{x}, \bar{y}_{Rob})}$$

และ $$\lambda_1 = R \cdot \frac{V(\bar{y}_{Rob}) - RCov(\bar{x}, \bar{y}_{Rob})}{V(\bar{y}_{Rob}) + R^2 V(\bar{x}) - 2RCov(\bar{x}, \bar{y}_{Rob})}$$

ความแปรปรวนที่เหมาะสมได้มาจากการแทนค่าที่เหมาะสมของ λ_0 และ λ_1 โดย

$$V(\bar{y}_{MU})_{opt} = \frac{\bar{Y}^2(S_{00}S_{11} - S_{10}^2)}{\bar{Y}^2 S_{11} - 2\bar{X}\bar{Y}S_{10} + \bar{X}^2 S_{00}}$$

$$= f\bar{Y}^2 C_X^2 (1 - \rho_{YX}^2)$$

2.1.5 การเลือกตัวอย่างสองเฟส (Two-Phase Sampling) และตัวประมาณค่าเฉลี่ยของประชากร

ในการประมาณค่าเฉลี่ยของตัวแปรที่สนใจศึกษา Y โดยใช้ตัวประมาณอัตราส่วน ตัวประมาณผลคูณ หรือตัวประมาณถดถอยเชิงเส้น ต้องทราบค่าเฉลี่ยประชากรของตัวแปรช่วย $\bar{X}$ ในบางครั้งไม่สามารถทราบค่า $\bar{X}$ ได้ แต่สามารถเก็บค่าข้อมูลของตัวแปรช่วย X และตัวแปรที่สนใจศึกษา Y ได้เมื่อทำการเลือกตัวอย่างมาแล้ว

ในบางสถานการณ์ตัวแปรที่สนใจศึกษา Y ยากต่อการเก็บข้อมูลหรือค่าใช้จ่ายในการเก็บข้อมูลมีราคาแพงมาก แต่ถ้ามีตัวแปรอื่นที่มีความสัมพันธ์กับตัวแปรที่สนใจศึกษาและสามารถเก็บข้อมูลได้ง่ายกว่าหรือมีค่าใช้จ่ายที่ถูกกว่า เรียกตัวแปรนี้ว่าตัวแปรช่วย X โดยทำการเก็บข้อมูลตัวแปรช่วย X จากตัวอย่างขนาดใหญ่กว่าข้อมูลจากตัวแปรที่สนใจศึกษา Y เรียกวิธีการ เลือกตัวอย่างนี้ว่า "การเลือกตัวอย่างสองเฟส" (Double sampling หรือ Two-phase sampling)
วิธีการเลือกตัวอย่างสองเฟสทำได้ดังนี้
เฟสที่ 1 ประชากรขนาด N หน่วย เลือกตัวอย่างขนาด $n^{(1)}$ หน่วย จากนั้นเก็บข้อมูลเฉพาะตัวแปรช่วย X เท่านั้น
เฟสที่ 2 จากตัวอย่างขนาด $n^{(1)}$ หน่วยในเฟสที่ 1 เลือกตัวอย่างขนาด $n^{(2)}$ หน่วย จากนั้นเก็บข้อมูลตัวแปรที่สนใจศึกษา Y และตัวแปรช่วย X

ตัวประมาณค่าเฉลี่ยของประชากรภายใต้การเลือกตัวอย่างสองเฟส

ตัวประมาณอัตราส่วน

กำหนดให้ y_i คือค่าสังเกตของตัวแปรที่สนใจศึกษาหน่วยที่ i
x_i คือค่าสังเกตของตัวแปรช่วย หน่วยที่ i
$n^{(1)}$ คือขนาดตัวอย่างของการเลือกตัวอย่างเฟสที่ 1
$n^{(2)}$ คือขนาดตัวอย่างของการเลือกตัวอย่างเฟสที่ 2

สำหรับเฟสที่ 2 นั้นหน่วยตัวอย่างแต่ละหน่วยจะมีค่าสังเกต x_i และ y_i ในขณะที่เฟสที่ 1 หน่วยตัวอย่างแต่ละหน่วยจะมีเฉพาะค่าสังเกต x_i

จากประชากรขนาด N เลือกตัวอย่างขนาด $n^{(1)}$ หน่วยโดยการเลือกตัวอย่างสุ่มแบบง่ายไม่คืนที่ จากนั้นเก็บข้อมูลเฉพาะตัวแปรช่วย x จากตัวอย่างขนาด $n^{(1)}$ หน่วย เลือกตัวอย่างขนาด $n^{(1)}$ หน่วย $\left(n^{(2)} < n^{(1)}\right)$ โดยการเลือกตัวอย่างสุ่มแบบง่ายไม่คืนที่ จากนั้นเก็บข้อมูลเฉพาะตัวแปรที่สนใจศึกษา y ดังนั้นในเฟสที่ 2 จะได้

$$\hat{R}^{(2)} = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i} = \frac{\sum_{i=1}^{n} y_i \Big/ n^{(2)}}{\sum_{i=1}^{n} x_i \Big/ n^{(2)}} = \frac{\bar{y}^{(2)}}{\bar{x}^{(2)}}$$

จากตัวอย่างขนาด $n^{(1)}$ หน่วย ในเฟสที่ 1 สามารถประมาณค่าเฉลี่ยของตัวแปรช่วย x ได้ดังนี้

$$\bar{x}^{(1)} = \frac{1}{n^{(1)}} \sum_{i=1}^{n^{(1)}} x_i$$

ดังนั้นตัวประมาณแบบอัตราส่วนเพื่อประมาณค่าเฉลี่ยของประชากร $\bar{X}$ คือ

$$\bar{y}_R^{(2)} = \hat{R}^{(2)} \bar{x}^{(1)}$$

โดยมีค่าความแปรปรวนดังนี้

$$V\left(\bar{y}_R^{(2)}\right) \approx \bar{Y}^2 \left[f_1 C_Y^2 + f_2 \left(C_Y^2 + C_X^2 - 2\rho_{YX} C_Y C_X \right) \right]$$

เมื่อ $f_1 = \frac{1}{n^{(1)}} - \frac{1}{N}$, $f_2 = \frac{1}{n^{(2)}} - \frac{1}{n^{(1)}}$, $C_Y^2 = \frac{S_y^2}{\bar{Y}^2}$, $C_X^2 = \frac{S_x^2}{\bar{X}^2}$, $\rho_{YX} = \frac{S_{yx}}{S_y S_x}$,

$S_y^2 = \frac{1}{(N-1)} \sum_{i=1}^{N} \left(y_i - \bar{Y}\right)^2$, $S_x^2 = \frac{1}{(N-1)} \sum_{i=1}^{N} \left(x_i - \bar{X}\right)^2$,

$S_{yx}^2 = \frac{1}{(N-1)} \sum_{i=1}^{N} \left(y_i - \bar{Y}\right)\left(x_i - \bar{X}\right)$

ตัวประมาณผลคูณ

ตัวประมาณค่าเฉลี่ยของประชากร $\bar{y}_P^{(2)} = \bar{y}^{(2)} \left(\frac{\bar{x}^{(2)}}{\bar{x}^{(1)}} \right)$

โดยมีค่าความแปรปรวนดังนี้

$$V\left(\bar{y}_P^{(2)}\right) \approx \bar{Y}^2 \left[f_1 C_Y^2 + f_2 \left(C_Y^2 + C_X^2 + 2\rho_{YX} C_Y C_X \right) \right]$$

2.1.6 ตัวประมาณไม่เอนเอียง (Unbiased Estimator)

ในการจะเลือกใช้ตัวประมาณเพื่อประมาณค่าพารามิเตอร์ θ นั้นจะต้องอาศัยการแจกแจงของตัวอย่างเป็นหลักในการพิจารณา โดยใช้คุณสมบัติของสถิติหลายอย่างประกอบกันในการเลือกตัวประมาณ ตัวประมาณที่ดีควรมีคุณสมบัติดังนี้ ได้แก่ ความไม่เอนเอียง (Unbiasedness) ความคงเส้นคงวา (Consistency) ความพอเพียง (Sufficiency) ความมีความแปรปรวนต่ำสุด (Minimum variance) ความมีประสิทธิภาพ (Efficiency) และความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุด (Minimum mean square error) ซึ่งในการวิจัยครั้งนี้ผู้วิจัยมุ่งเน้นศึกษาตัวประมาณไม่เอนเอียง

ตัวประมาณไม่เอนเอียง (Unbiased Estimator) ถ้า $\hat{\theta}$ เป็นตัวประมาณของพารามิเตอร์ θ ใด ๆ $\hat{\theta}$ จะเป็นตัวประมาณที่ไม่เอนเอียง (Unbiased estimator) ก็ต่อเมื่อ $E\left(\hat{\theta}\right) = \theta$ และ $\hat{\theta}$ จะเป็นตัวประมาณที่เอนเอียง (Biased estimator) ก็ต่อเมื่อ $E\left(\hat{\theta}\right) \neq \theta$ เพราะฉะนั้น ความเอนเอียงคือ $Bias\left(\hat{\theta}\right) = E\left(\hat{\theta}\right) - \theta$ (Singh & Pal, 2015)

ถ้าพิจารณาค่าคาดหวังของ $\hat{\theta}$ จะเห็นได้ว่าถ้าค่าเฉลี่ยของการแจกแจงจากตัวอย่างของ $\hat{\theta}$ มีค่าเท่ากับพารามิเตอร์ θ พอดี แสดงว่าจุดกลางของการแจกแจงของตัวประมาณอยู่ ณ จุดที่เป็นค่าจริง ดังนั้น คุณสมบัติประการหนึ่งของตัวประมาณที่ควรพิจารณา คือ ความไม่เอนเอียงของตัวประมาณ

2.1.7 การเปรียบเทียบประสิทธิภาพของตัวประมาณ (Comparison of Efficiency)

ในการวิจัยครั้งนี้ผู้วิจัยเปรียบประสิทธิภาพของตัวประมาณด้วยความแปรปรวนต่ำสุด (Minimum Variance) ถ้ามีตัวประมาณที่ไม่เอนเอียง 2 ตัว ที่ใช้ประมาณค่าพารามิเตอร์ตัวเดียวกัน หาประสิทธิภาพของตัวประมาณได้ โดยการนำความแปรปรวนของตัวประมาณหนึ่งไปลบความแปรปรวนของตัวประมาณอีกตัวหนึ่ง โดยทั่วไปเรามักจะเปรียบเทียบประสิทธิภาพของตัวประมาณที่สนใจกับตัวประมาณที่มีความแปรปรวนต่ำสุด โดยทั่วไป ตัวประมาณเป็นตัวประมาณที่ดี ควรที่จะมีความคลาดเคลื่อนต่ำ ซึ่งความคลาดเคลื่อนของตัวประมาณพิจารณาจากความแปรปรวนหรือส่วนเบี่ยงเบนมาตรฐานจึงอาจใช้ความแปรปรวนของตัวประมาณในการหาประสิทธิภาพของตัวประมาณได้ ในบางครั้ง อาจจะมีตัวประมาณที่ไม่เอนเอียงและคงเส้นคงวาได้หลายตัว ในการเลือกตัวใดตัวหนึ่งนั้นจะอาศัยคุณสมบัติความมีประสิทธิภาพสัมพัทธ์มาตัดสินใจเลือก

เปรียบเทียบประสิทธิภาพของตัวประมาณโดยพิจารณาจากค่าประมาณความแปรปรวนของตัวประมาณ $\hat{V}(\bar{y}_1)-\hat{V}(\bar{y}_2)$

โดย $\hat{V}(\bar{y}_1)-\hat{V}(\bar{y}_2)>0$ แสดงว่าตัวประมาณค่าเฉลี่ยประชากร $\bar{y}_2$ มีประสิทธิภาพดีกว่า $\bar{y}_1$

$\hat{V}(\bar{y}_1)-\hat{V}(\bar{y}_2)<0$ แสดงว่าตัวประมาณค่าเฉลี่ยประชากร $\bar{y}_1$ มีประสิทธิภาพดีกว่า $\bar{y}_2$

2.2 งานวิจัยที่เกี่ยวข้อง

2.2.1 งานวิจัยภายในประเทศ

วิชชญาพร สิงห์นวล และ จิราวัลย์ จิตรถเวช (2559) ได้ศึกษาเรื่อง การเปรียบเทียบประสิทธิภาพของตัวประมาณค่าต่างๆกับตัวประมาณค่าของซิลส์ที่ปรับปรุงใหม่ โดยงานวิจัยนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของตัวประมาณค่าเฉลี่ยของประชากรที่ใช้การสุ่มตัวอย่างแบบง่าย ซึ่งการศึกษานี้ใช้ตัวประมาณ 7 ตัว ได้แก่ ตัวประมาณค่าแบบอัตราส่วน ตัวประมาณค่าแบบการถดถอยเชิงเส้น 2 วิธีที่ใช้ตัวประมาณค่าของซิลล์มาปรับปรุงตัวประมาณค่าของซิลส์แบบดั้งเดิม ตัวประมาณค่าของซิโซเดียและไวว์ดี (Sisodia and Dwivedi) ตัวประมาณค่าของโคยุนคุและคาดิลาร์ (Koyuncu and Kadilar) และตัวประมาณค่าของแชบเบียร์และคุปตะ (Shabbir and Gupta) โดยใช้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) เป็นเกณฑ์ในการเปรียบเทียบของตัวประมาณค่า ข้อมูลที่ใช้เป็นข้อมูลค่าเฉลี่ยจำนวนอาจารย์ต่อสถานศึกษาในเขตกรุงเทพมหานครและเขตปริมณฑลจำนวน 1,009 โรงเรียน ผลการศึกษพบว่าตัวประมาณค่าเฉลี่ยโดยใช้วิธีการถดถอยเชิงเส้นทั้ง 2 ตัว ที่ใช้ตัวประมาณค่าของซิลล์มาปรับปรุงมีประสิทธิภาพดีกว่าตัวประมาณค่าแบบอื่นสอดคล้องกับผลงานวิจัยทางทฤษฎีที่ได้ศึกษาไว้แล้ว

จิราภรณ์ ธรรมสาโรช (2560) ได้ศึกษาเรื่อง ตัวประมาณอัตราส่วนร่วมกับผลคูณแบบปรับปรุงภายใต้การเลือกตัวอย่างสองเฟส วัตถุประสงค์ของการศึกษาคือเสนอตัวประมาณอัตราส่วนร่วมกับผลคูณสำหรับการประมาณค่าเฉลี่ยของประชากรภายใต้การเลือกตัวอย่างสองเฟสโดยใช้ตัวแปรช่วยสองตัวแปร ตัวประมาณที่นำเสนอปรับปรุงตัวประมาณของ Singh และคณะ ในปี 2005 และตัวประมาณของ Chanu และ Singh ในปี 2014 พร้อมทั้งหาความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณที่นำเสนอ ทำการเปรียบเทียบประสิทธิภาพของตัวประมาณที่นำเสนอกับตัวประมาณที่ไม่เอนเอียง ตัวประมาณของ Singh และคณะ ในปี 2005 และตัวประมาณของ Chanu และ Singh ในปี 2014 โดยใช้ความคลาดเคลื่อนกำลังสองเฉลี่ย ผลการศึกษาพบว่าตัวประมาณที่นำเสนอมีประสิทธิภาพมากกว่าตัวประมาณที่ไม่เอนเอียง ตัวประมาณของ Singh และคณะ ในปี 2005 และตัวประมาณของ Chanu และ Singh ในปี 2014 ภายใต้การศึกษาจากข้อมูลจริง

ณภัทณ์จันทร์ ด่านสวัสดิ์ (2563) ได้ศึกษาเรื่อง ตัวประมาณอัตราส่วนร่วมกับผลคูณแบบเลขชี้กำลัง สำหรับค่าเฉลี่ยประชากรในการเลือกตัวอย่างสุ่มแบบง่าย โดยใช้สารสนเทศจากตัวแปรช่วย งานวิจัยฉบับนี้นำเสนอตัวประมาณอัตราส่วนร่วมกับผลคูณแบบเลขชี้กำลังของค่าเฉลี่ยประชากรโดยใช้สารสนเทศจากตัวแปรช่วยภายใต้การเลือกตัวอย่างสุ่มแบบง่าย ซึ่งผู้วิจัยได้พัฒนาตัวประมาณนี้ขึ้นมาจากตัวประมาณที่นำเสนอโดย Singh and Pal (2015) นอกจากนี้สมบัติที่สำคัญบางประการของตัวประมาณอาทิเช่น ความเอนเอียง (Bias) ค่าคลาดเคลื่อน กำลังสองเฉลี่ย (Mean Squared Error: MSE) และค่าคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุด (Minimum Mean Squared Error: MMSE) จะถูกนำเสนอ ในขณะที่การเปรียบเทียบประสิทธิภาพของตัวประมาณที่นำเสนอกับตัวประมาณที่เกี่ยวข้องจะแสดงการคำนวณเชิงตัวเลขเพื่อสนับสนุนผลทางทฤษฎี ผลการวิจัยพบว่า ตัวประมาณอัตราส่วนร่วมกับผลคูณแบบเลขชี้กำลัง ที่นำเสนอมีประสิทธิภาพดีกว่าตัวประมาณอื่น ๆ ที่เกี่ยวข้อง

2.2. งานวิจัยต่างประเทศ

Vishwakarma และ Singh (2016) ได้ศึกษาเรื่อง การปรับปรุงตัวประมาณอัตราส่วนและตัวประมาณผลคูณเพื่อประมาณค่าเฉลี่ยของประชากรจำกัด ในการเลือกตัวอย่างแบบง่ายไม่คืนที่ โดยผู้วิจัยกล่าวถึงการปรับปรุงตัวประมาณอัตราส่วนและตัวประมาณผลคูณ เพื่อให้การประมาณค่าเฉลี่ยของประชากรมีประสิทธิภาพมากขึ้น ทำการศึกษาข้อมูลการใช้ตัวแปรช่วย ในการเลือกตัวอย่างแบบง่ายไม่คืนที่ แสดงความเอนเอียงและความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณภายใต้การประมาณอันดับหนึ่งการศึกษาเชิงทฤษฎีและเชิงตัวเลข มีการพิสูจน์แล้วว่าตัวประมาณที่นำเสนอนี้มีประสิทธิภาพมากกว่าตัวประมาณแบบอื่น ๆ ที่มีอยู่แล้ว

Mahanty และ Mishra (2020) ได้ศึกษาเรื่อง ตัวประมาณที่ไม่เอนเอียงของค่าเฉลี่ยประชากรแบบจำกัดโดยใช้สารสนเทศของตัวแปรช่วย โดยผู้วิจัยสร้างตัวประมาณที่ไม่เอนเอียงขึ้นโดยใช้ผลรวมเชิงเส้นของตัวประมาณตัวแปรที่ศึกษาและค่าเฉลี่ยต่อหน่วยของตัวแปรช่วยภายใต้การสุ่มตัวอย่างอย่างง่ายแบบไม่คืนที่ เปรียบเทียบประสิทธิภาพของตัวประมาณภายใต้สภาวะที่เหมาะสมที่สุดกับตัวประมาณค่าเฉลี่ยต่อหน่วย ตัวประมาณอัตราส่วนที่ไม่เอนเอง ตัวประมาณผลคูณที่ไม่เอนเอียง และตัวประมาณการถดถอย ทั้งในทางทฤษฎีและเชิงตัวเลข ผลการวิจัยพบว่า ตัวประมาณที่ผู้วิจัยเสนอ มีประสิทธิภาพสูงกว่าตัวประมาณอื่นที่ศึกษา

Vishwakarma และ Zeeshan (2021) ได้ศึกษาเรื่อง ตัวประมาณอัตราส่วนร่วมกับตัวประมาณผลคูณสำหรับค่าเฉลี่ยของประชากรจำกัดภายใต้แผนการเลือกตัวอย่างแบบสองเฟสโดยวิธีการลดค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Square Error: MSE) ของตัวประมาณค่าที่ผู้วิจัยเสนอเทียบกับค่า MSE ของตัวประมาณการถดถอยเชิงเส้นภายใต้รูปแบบการสุ่มตัวอย่างแบบสองตอนได้รับการพัฒนาขึ้น ตัวประมาณได้รับการพัฒนาเพื่อประเมินค่าเฉลี่ยของตัวแปรภายใต้การศึกษาโดยใช้ตัวแปรช่วย (ซึ่งไม่เป็นที่รู้จัก แต่สามารถเข้าถึงได้อย่างสะดวกและประหยัด) สมการความคลาดเคลื่อนกำลังสองเฉลี่ยจะได้รับสำหรับตัวประมาณค่าที่เสนอนอกจากนี้ยังได้ขนาดตัวอย่างที่เหมาะสมภายใต้ฟังก์ชันต้นทุนที่กำหนด มีการศึกษาเปรียบเทียบเพื่อกำหนดเงื่อนไขซึ่งตัวประมาณค่าที่พัฒนาขึ้นมีประสิทธิภาพมากกว่าตัวประมาณค่าอื่นที่มีความแปลกใหม่ การศึกษาเชิงประจักษ์ยังดำเนินการเพื่อเสริมการอ้างว่าตัวประมาณค่าที่พัฒนาขึ้นนั้นมีประสิทธิภาพมากกว่า

บทที่ 3
วิธีดำเนินการวิจัย

งานวิจัยเรื่อง “ตัวประมาณไม่เอนเอียงโดยการใช้สารสนเทศของตัวแปรช่วยสำหรับค่าเฉลี่ยประชากรจำกัดภายใต้การเลือกตัวอย่างสองเฟส” มีวิธีการดำเนินการวิจัย ดังต่อไปนี้

3.1 ศึกษาและพัฒนาตัวประมาณไม่เอนเอียงโดยใช้สารสนเทศของตัวแปรช่วย
3.2 ศึกษาตัวประมาณโดยใช้ข้อมูลจำลอง
3.3 การเปรียบประสิทธิภาพ
3.4 แผนภาพขั้นตอนวิธีการดำเนินงาน

3.1 ศึกษาและพัฒนาตัวประมาณไม่เอนเอียงโดยใช้สารสนเทศของตัวแปรช่วย

ผู้วิจัยศึกษาและพัฒนาทฤษฎีเกี่ยวกับตัวประมาณไม่เอนเอียงโดยใช้สารสนเทศของตัวแปรช่วยสำหรับค่าเฉลี่ยประชากรจำกัดภายใต้การตัวอย่างสองเฟส โดยใช้แนวคิดของ Tin (1965) Robson (1957) และ Mahanty & Mishra (2020)

โดย Tin (1965) เสนอตัวประมาณอัตราส่วนไม่เอนเอียงสำหรับการเลือกตัวอย่างแบบ SRSWOR มีรูปแบบดังนี้

$$\bar{y}_{Tin} = \bar{y}\frac{\bar{X}}{\bar{x}}\left[1 + f\left(\frac{s_{yx}}{\bar{y}\,\bar{x}} - \frac{s_x^2}{\bar{x}^2}\right)\right]$$

โดยมีค่าความแปรปรวนคือ

$$V\left(\bar{y}_{Tin}\right) \approx f\bar{Y}^2\left[C_Y^2 + C_X^2 - 2\rho_{YX}C_YC_X\right]$$

Robson (1957) เสนอตัวประมาณผลคูณไม่เอนเอียงสำหรับการเลือกตัวอย่างแบบ SRSWOR มีรูปแบบดังนี้

$$\bar{y}_{Rob} = \frac{\bar{y}\,\bar{x}}{\bar{X}} - f\frac{s_{yx}}{\bar{X}}$$

โดยมีค่าความแปรปรวนคือ

$$V\left(\bar{y}_{Rob}\right) \approx f\bar{Y}^2\left[C_Y^2 + C_X^2 + 2\rho_{YX}C_YC_X\right]$$

Mahanty & Mishra (2020) เสนอตัวประมาณไม่เอนเอียงสำหรับการเลือกตัวอย่างแบบ SRSWOR มีรูปแบบดังนี้

$$\bar{y}_{MU} = \lambda_0\left(\bar{y}\frac{\bar{x}}{\bar{X}} - f\frac{s_{xy}}{\bar{X}}\right) + \lambda_1\bar{x}$$

โดยมีค่าความแปรปรวนคือ

$$V\left(\bar{y}_{MU}\right)_{opt} \approx f\bar{Y}^2C_X^2\left(1 - \rho_{YX}^2\right)$$

3.2 ศึกษาตัวประมาณโดยใช้ข้อมูลจำลอง

ผู้วิจัยใช้ข้อมูลจำลองผ่านการใช้โปรแกรม RStudio ในการจำลองข้อมูลสร้างกลุ่มประชากรขึ้นมา เพื่อทดสอบทฤษฎีของตัวประมาณภายใต้แนวคิดของ Tin (1965) Robson (1957) และ Mahanty & Mishra (2020) และพัฒนาตัวประมาณไม่เอนเอียงโดยการใช้สารสนเทศของตัวแปรช่วยสำหรับค่าเฉลี่ยประชากรจำกัด ภายใต้การเลือกตัวอย่างสองเฟส โดยสร้างประชากรให้มีการแจกแจงปรกติ และกำหนดความสัมพันธ์กันในระดับต่าง ๆ โดย

สร้างประชากรขนาด $N=1,000$ หน่วย โดย $Y \sim N\left(100, 20^2\right)$ และ $X \sim N\left(500, 100^2\right)$

กรณี Y และ X มีความสัมพันธ์เชิงเส้นตรงในทิศทางเดียวกัน กำหนดค่าความสัมพันธ์ของประชากร 3 ระดับ คือ

1 ประชากรมีความสัมพันธ์กันสูง $\rho_{YX}=0.8$
2 ประชากรมีความสัมพันธ์กันปานกลาง $\rho_{YX}=0.6$
3 ประชากรมีความสัมพันธ์กันน้อย $\rho_{YX}=0.3$

กรณี Y และ X มีความสัมพันธ์เชิงเส้นตรงในทิศทางตรงกันข้าม กำหนดค่าความสัมพันธ์ของประชากร 3 ระดับ คือ

1 ประชากรมีความสัมพันธ์กันสูง $\rho_{YX}=-0.8$
2 ประชากรมีความสัมพันธ์กันปานกลาง $\rho_{YX}=-0.6$
3 ประชากรมีความสัมพันธ์กันน้อย $\rho_{YX}=-0.3$

ทำการสุ่มข้อมูลภายใต้การเลือกตัวอย่างสองเฟส 8 รูปแบบ ดังนี้

N	$n^{(1)}$	ร้อยละ$^{(1)}$	$n^{(2)}$	ร้อยละ$^{(2)}$
1,000	50	5	10	20
1,000	100	10	20	20
1,000	100	10	50	50
1,000	200	20	50	25
1,000	200	20	100	50
1,000	500	50	100	20
1,000	500	50	200	40
1,000	500	50	300	60

โดย ร้อยละ$^{(1)}$ คือ สัดส่วนของการเลือกตัวอย่างเฟสที่ 1 จากประชากร
ร้อยละ$^{(2)}$ คือ สัดส่วนของการเลือกตัวอย่างเฟสที่ 2 จากตัวอย่างเฟสที่ 1

3.3 การเปรียบเทียบประสิทธิภาพ

ในการวิจัยครั้งนี้ผู้วิจัยจะเปรียบเทียบประสิทธิภาพของตัวประมาณที่ศึกษากับตัวประมาณอัตราส่วนและตัวประมาณผลคูณแบบทั่วไป แบ่งออกเป็น 2 รูปแบบ

3.3.1 เปรียบเทียบประสิทธิภาพของตัวประมาณค่าในเชิงทฤษฎี

เปรียบเทียบประสิทธิภาพของตัวประมาณค่า โดยใช้ค่าความแปรปรวนต่ำสุด (Minimum Variance) ในเชิงทฤษฎี

3.3.2 เปรียบเทียบประสิทธิภาพของตัวประมาณค่าจากการจำลองข้อมูล

ในการจำลองข้อมูลของแต่ละตัวประมาณ จะดำเนินการทำซ้ำทั้งหมด $5,000$ รอบ

คำนวณค่า $$\bar{y}_i = \frac{\sum_{k=1}^{5,000} \bar{y}_k}{5,000}$$ เมื่อ i คือ ตัวประมาณที่ศึกษาตัวที่ i

และ $$Bi\hat{a}s(\bar{y}_i) = \frac{\sum_{k=1}^{5,000} Bi\hat{a}s(\bar{y}_k)}{5,000}$$

และ $$\hat{V}(\bar{y}_i) = \frac{\sum_{k=1}^{5,000} \hat{V}(\bar{y}_k)}{5,000}$$

เปรียบเทียบประสิทธิภาพของตัวประมาณโดยพิจารณาจากค่าประมาณความแปรปรวนของตัวประมาณ $\hat{V}(\bar{y}_1) - \hat{V}(\bar{y}_2)$

โดย $\hat{V}(\bar{y}_1) - \hat{V}(\bar{y}_2) > 0$ แสดงว่าตัวประมาณค่าเฉลี่ยประชากร $\bar{y}_2$ มีประสิทธิภาพดีกว่า $\bar{y}_1$

$\hat{V}(\bar{y}_1) - \hat{V}(\bar{y}_2) < 0$ แสดงว่าตัวประมาณค่าเฉลี่ยประชากร $\bar{y}_1$ มีประสิทธิภาพดีกว่า $\bar{y}_2$

3.4 แผนภาพขั้นตอนวิธีการดำเนินงาน

สร้างประชากรขนาด $N = 1,000$ หน่วย โดย $Y \sim N\left(100, 20^2\right)$ และ $X \sim N\left(500, 100^2\right)$

กำหนดค่าความสัมพันธ์ของประชากร 3 ระดับ

ในกรณี Y และ X มีความสัมพันธ์กันทางบวกและทางลบ

1 ประชากรมีความสัมพันธ์กันสูง $\rho_{YX} = 0.8$ และ $\rho_{YX} = -0.8$

2 ประชากรมีความสัมพันธ์กันปานกลาง $\rho_{YX} = 0.6$ และ $\rho_{YX} = -0.6$

3 ประชากรมีความสัมพันธ์กันน้อย $\rho_{YX} = 0.3$ และ $\rho_{YX} = -0.3$

↓

จากประชากรที่มีความสัมพันธ์กันทั้ง 3 ระดับ

จะทำการสุ่มข้อมูลภายใต้การเลือกตัวอย่างสองเฟส 8 รูปแบบ ดังนี้

1. $n^{(1)} = 50$ และ $n^{(2)} = 10$
2. $n^{(1)} = 100$ และ $n^{(2)} = 20$
3. $n^{(1)} = 100$ และ $n^{(2)} = 50$
4. $n^{(1)} = 200$ และ $n^{(2)} = 50$
5. $n^{(1)} = 200$ และ $n^{(2)} = 100$
6. $n^{(1)} = 500$ และ $n^{(2)} = 100$
7. $n^{(1)} = 500$ และ $n^{(2)} = 200$
8. $n^{(1)} = 500$ และ $n^{(2)} = 300$

↓

ประมาณค่าของประชากรโดยใช้

กรณี $\rho_{YX} < 0$	กรณี $\rho_{YX} > 0$
ตัวประมาณค่าเฉลี่ย ได้แก่	ตัวประมาณค่าเฉลี่ย ได้แก่
$\bar{y}_P^{(2)}, \bar{y}_{Rob}^{(2)}, \bar{y}_{MU}^{(2)}$	$\bar{y}_R^{(2)}, \bar{y}_{Tin}^{(2)}, \bar{y}_{MU}^{(2)}$
ค่าประมาณความแปรปวน ได้แก่	ค่าประมาณความแปรปวน ได้แก่
$\hat{V}\left(\bar{y}_P^{(2)}\right), \hat{V}\left(\bar{y}_{Rob}^{(2)}\right), \hat{V}\left(\bar{y}_{MU}^{(2)}\right)$	$\hat{V}\left(\bar{y}_R^{(2)}\right), \hat{V}\left(\bar{y}_{Tin}^{(2)}\right), \hat{V}\left(\bar{y}_{MU}^{(2)}\right)$

สุ่มข้อมูลทำซ้ำจากประชากรทั้งหมด 5,000 รอบ

คำนวณค่า $\bar{y}_i = \dfrac{\sum_{k=1}^{5,000} \bar{y}_k}{5,000}$

และ $Bi\hat{a}s(\bar{y}_i) = \dfrac{\sum_{k=1}^{5,000} Bi\hat{a}s(\bar{y}_k)}{5,000}$

และเปรียบเทียบประสิทธิภาพของตัวประมาณ โดย

$$\hat{V}(\bar{y}_i) = \frac{\sum_{k=1}^{5,000} \hat{V}(\bar{y}_k)}{5,000}$$

บทที่ 4
ผลการวิจัยและการอภิปราย

ในการวิจัยตัวประมาณไม่เอนเอียงโดยการใช้สารสนเทศของตัวแปรช่วยสำหรับค่าเฉลี่ยประชากรจำกัดภายใต้การเลือกตัวอย่างแบบสองเฟส โดยเปรียบเทียบประสิทธิภาพของตัวประมาณค่าในเชิงทฤษฎี และเปรียบเทียบประสิทธิภาพของตัวประมาณค่าจากการจำลองข้อมูล ซึ่งมีรายละเอียดดังต่อไปนี้

4.1 ตัวประมาณไม่เอนเอียง ภายใต้การเลือกตัวอย่างสองเฟส (Proposed Unbiased Estimator under Two-Phase Sampling)

4.2 เปรียบเทียบประสิทธิภาพของตัวประมาณ (Comparison of Efficiency)

4.3 การอธิบายเชิงตัวเลข (Numerically Illustrations)

4.1 ตัวประมาณไม่เอนเอียง ภายใต้การเลือกตัวอย่างสองเฟส (Proposed Unbiased Estimator under Two-Phase Sampling)

ถ้าหากว่าตัวแปร X มีความสัมพันธ์ในทิศทางเดียวกันกับตัวแปรที่ต้องการศึกษา Y ในระดับสูงจะเรียกตัวประมาณดังกล่าวว่า ตัวประมาณอัตราส่วน (Ratio Estimator)

ในทางกลับกัน ถ้าหากว่าตัวแปร X มีความสัมพันธ์ในทิศทางตรงกันข้ามกับตัวแปรที่ต้องการศึกษา Y ในระดับสูงจะเรียกตัวประมาณดังกล่าวว่า ตัวประมาณผลคูณ (Product Estimator)

ในการวิจัยครั้งนี้ผู้วิจัยจึงสนใจศึกษาทฤษฎีเกี่ยวกับตัวประมาณอัตราส่วนไม่เอนเอียงและตัวประมาณผลคูณไม่เอนเอียงดังนี้ ตัวประมาณอัตราส่วนที่พัฒนาโดย Tin (1965) ตัวประมาณผลคูณของ Robson (1957) และ ตัวประมาณไม่เอนเอียงของ Mahanty และ Mishra (2020) ซึ่งเป็นตัวประมาณภายใต้การเลือกตัวอย่างแบบง่ายไม่คืนที่ (Simple Random Sampling without Replacement: SRSWOR)

ตัวประมาณอัตราส่วนที่เสนอโดย Tin (1965)

ได้นำเสนอตัวประมาณอัตราส่วนไม่เอนเอียงในรูปแบบใหม่ ที่มีค่าความแปรปรวนต่ำ เพื่อใช้ในการประมาณค่าเฉลี่ยของประชากรจำกัด โดย

$$\bar{y}_{Tin} = \bar{y}\frac{\bar{X}}{\bar{x}}\left[1+f\left(\frac{s_{yx}}{\bar{y}\,\bar{x}}-\frac{s_x^2}{\bar{x}^2}\right)\right]$$

โดยมีค่าความแปรปรวนคือ

$$V(\bar{y}_{Tin}) \approx f\bar{Y}^2\left[C_Y^2+C_X^2-2\rho_{YX}C_YC_X\right]$$

ตัวประมาณผลคูณที่เสนอโดย Robson (1957)

ได้นำเสนอตัวประมาณผลคูณไม่เอนเอียงในรูปแบบใหม่ ที่มีค่าความคลาดเคลื่อนต่ำ เพื่อใช้ในการประมาณค่าเฉลี่ยของประชากรจำกัด โดย

$$\bar{y}_{Rob} = \frac{\bar{y}\,\bar{x}}{\bar{X}} - f\,\frac{s_{yx}}{\bar{X}}$$

โดยมีค่าความแปรปรวนคือ

$$V\left(\bar{y}_{Rob}\right) \approx f\bar{Y}^2\left[C_Y^2 + C_X^2 + 2\rho_{YX}C_YC_X\right]$$

ตัวประมาณที่เสนอโดย Mahanty & Mishra (2020)

ได้สร้างตัวประมาณที่ไม่เอนเอียงขึ้นโดยใช้ผลรวมเชิงเส้นของตัวประมาณตัวแปรที่ศึกษาและค่าเฉลี่ยต่อหน่วยของตัวแปรช่วยภายใต้การสุ่มตัวอย่างอย่างง่ายแบบไม่คืนที่ เมื่อพิจารณาผลรวมเชิงเส้นของตัวประมาณที่ไม่เอนเอียงของร็อบสันและตัวประมาณค่าเฉลี่ยต่อหน่วยของตัวแปรช่วย โดย

$$\bar{y}_{MU} = \lambda_0\left(\bar{y}\,\frac{\bar{x}}{\bar{X}} - f\,\frac{s_{xy}}{\bar{X}}\right) + \lambda_1\bar{x}$$

โดยมีค่าความแปรปรวนคือ

$$V\left(\bar{y}_{MU}\right)_{opt} \approx f\bar{Y}^2C_X^2\left(1-\rho_{YX}^2\right)$$

ดังนั้นผู้วิจัยจึงพัฒนาตัวประมาณเหล่านี้ในทางทฤษฎี ภายใต้การเลือกตัวอย่างสองเฟส (Two-Phase Sampling)

สมมติ X เป็นตัวแปรมีความสัมพันธ์สูงกับตัวแปรที่สนใจศึกษา Y

กำหนดให้ $\bar{x}^{(1)}$ คือค่าเฉลี่ยของตัวแปรช่วยในการเลือกตัวอย่างเฟสที่ 1

$\bar{x}^{(2)}$ คือค่าเฉลี่ยของตัวแปรช่วยในการเลือกตัวอย่างเฟสที่ 2

$\bar{y}^{(2)}$ คือค่าเฉลี่ยของตัวแปรที่สนใจศึกษาในการเลือกตัวอย่างเฟสที่ 2

$n^{(1)}$ คือขนาดตัวอย่างของการเลือกตัวอย่างเฟสที่ 1

$n^{(2)}$ คือขนาดตัวอย่างของการเลือกตัวอย่างเฟสที่ 2

จากประชากรขนาด N เลือกตัวอย่างขนาด $n^{(1)}$ หน่วยโดยการเลือกตัวอย่างสุ่มแบบง่ายไม่คืนที่ จากนั้นเก็บข้อมูลเฉพาะตัวแปรช่วย x จากตัวอย่างขนาด $n^{(1)}$ หน่วย เลือกตัวอย่างขนาด $n^{(2)}$ หน่วย $\left(n^{(2)} < n^{(1)}\right)$ โดยการเลือกตัวอย่างสุ่มแบบง่ายไม่คืนที่ จากนั้นเก็บข้อมูลเฉพาะตัวแปรที่สนใจศึกษา y

กำหนดให้ Z_i คือตัวแปรสุ่มหน่วยที่ i

$$Z_i = \begin{cases} 1 & \text{ถ้าหน่วยที่ } i \text{ ตกอยู่ในเฟสที่ 1} \\ 0 & \text{ถ้าหน่วยที่ } i \text{ ไม่ได้ตกอยู่ในเฟสที่ 1} \end{cases}$$

ความน่าจะเป็นที่หน่วยตัวอย่างจะอยู่ในเฟสที่ 2 ขึ้นอยู่กับว่าหน่วยตัวอย่างนั้นอยู่ในเฟสที่ 1 หรือไม่ และอาจขึ้นอยู่กับสารสนเทศข้อมูลเสริมที่รวมในตัวเฟสที่ 1 ซึ่งสามารถเขียนแทนสัญลักษณ์คือ $P\left(D_i = 1 \mid \mathbb{Z}\right)$ โดยที่ $\mathbb{Z}$ คือเวกเตอร์ของ $\mathbb{Z} = \left[Z_1 \quad Z_2 \quad \dots \quad Z_N\right]$

เมื่อ $D_i = \begin{cases} 1 & \text{ถ้าหน่วยที่ } i \text{ ตกอยู่ในเฟสที่ 2} \\ 0 & \text{ถ้าหน่วยที่ } i \text{ ไม่ได้ตกอยู่ในเฟสที่ 2} \end{cases}$ (Lohr, 2022)

4.1.1 ตัวประมาณอัตราส่วนที่เสนอโดย Tin (1965)

เมื่อพิจารณาตัวประมาณอัตราส่วนไม่เอนเอียงของ Tin (1965) ภายใต้การเลือกตัวอย่างแบบสองเฟสแล้ว จะได้

$$\bar{y}_{Tin}^{(2)} = \bar{y}^{(2)}\frac{\bar{x}^{(1)}}{\bar{x}^{(2)}}\left[1+f_2\left(\frac{s_{yx}^{(2)}}{\bar{x}^{(2)}\bar{y}^{(2)}}-\frac{\left(s_x^{(2)}\right)^2}{\left(\bar{x}^{(2)}\right)^2}\right)\right]$$

โดยมีค่าความแปรปรวนดังนี้

พิจารณา $V\left(\bar{y}_{Tin}^{(2)}\right) \approx V_1\left\{E_2\left(\bar{y}_{Tin}^{(2)} \mid \mathbb{Z}\right)\right\} + E_1\left\{V_2\left(\bar{y}_{Tin}^{(2)} \mid \mathbb{Z}\right)\right\}$

จาก $\bar{y}_{Tin}^{(2)}$ เป็นตัวประมาณอัตราส่วนไม่เอนเอียง ดังนั้น $E_2\left(\bar{y}_{Tin}^{(2)} \mid \mathbb{Z}\right) \to \bar{y}^{(1)}$

จะได้ $V_1\left\{E_2\left(\bar{y}_{Tin}^{(2)} \mid \mathbb{Z}\right)\right\} = V_1\left(\bar{y}^{(1)}\right)$

พิจารณา $E_1\left\{V_2\left(\bar{y}_{Tin}^{(2)} \mid \mathbb{Z}\right)\right\}$

โดยที่ $V_2\left(\bar{y}_{Tin}^{(2)} \mid \mathbb{Z}\right) = f_2\bar{Y}^2\left(C_Y^2 + C_X^2 - 2\rho_{YX}C_YC_X\right)$

ดังนั้น

$$V\left(\bar{y}_{Tin}^{(2)}\right) \approx V_1\left(\bar{y}^{(1)}\right) + V_2\left(\bar{y}_{Tin}^{(2)}\right)$$

$$\approx \left(f_1\bar{Y}^2C_Y^2\right) + \left(f_2\bar{Y}^2\left(C_Y^2 + C_X^2 - 2\rho_{YX}C_YC_X\right)\right)$$

$$V\left(\bar{y}_{Tin}^{(2)}\right) \approx \bar{Y}^2\left(f_1C_Y^2 + f_2\left(C_Y^2 + C_X^2 - 2\rho_{YX}C_YC_X\right)\right)$$

เมื่อค่า $f_1 = \frac{1}{n^{(1)}} - \frac{1}{N}$, $f_2 = \frac{1}{n^{(2)}} - \frac{1}{n^{(1)}}$, $C_Y^2 = \frac{S_y^2}{\bar{Y}^2}$, $C_X^2 = \frac{S_x^2}{\bar{X}^2}$, $\rho_{YX} = \frac{S_{yx}}{S_yS_x}$,

$$S_y^2 = \frac{1}{(N-1)}\sum_{i=1}^{N}\left(y_i - \bar{Y}\right)^2,\ S_x^2 = \frac{1}{(N-1)}\sum_{i=1}^{N}\left(x_i - \bar{X}\right)^2,$$

$$S_{yx}^2 = \frac{1}{(N-1)}\sum_{i=1}^{N}\left(y_i - \bar{Y}\right)\left(x_i - \bar{X}\right)$$

เมื่อ f_1 คือ เศษส่วนการเลือกตัวอย่างในเฟสที่ 1

f_2 คือ เศษส่วนการเลือกตัวอย่างในเฟสที่ 2

C_Y คือ สัมประสิทธิ์ความแปรผันของประชากรของตัวแปร Y

C_X คือ สัมประสิทธิ์ความแปรผันของประชากรของตัวแปร X

ρ_{YX} คือ สัมประสิทธิ์สหสัมพันธ์ของตัวแปร X และ ตัวแปร Y

ตัวประมาณของ $V\left(\bar{y}_{Tin}^{(2)}\right)$ คือ

$$\hat{V}\left(\bar{y}_{Tin}^{(2)}\right) \approx \left(\bar{y}^{(2)}\right)^2\left(f_1\left(c_y^{(2)}\right)^2 + f_2\left(\left(c_y^{(2)}\right)^2 + \left(c_x^{(2)}\right)^2 - 2r_{yx}^{(2)}c_y^{(2)}c_x^{(2)}\right)\right)$$

เมื่อค่า $f_1 = \frac{1}{n^{(1)}} - \frac{1}{N}$, $f_2 = \frac{1}{n^{(2)}} - \frac{1}{n^{(1)}}$, $\left(c_y^{(2)}\right)^2 = \frac{\left(s_y^{(2)}\right)^2}{\left(\bar{y}^{(2)}\right)^2}$, $\left(c_x^{(2)}\right)^2 = \frac{\left(s_x^{(2)}\right)^2}{\left(\bar{x}^{(2)}\right)^2}$,

$$r_{yx}^{(2)}=\frac{s_{yx}^{(2)}}{s_y^{(2)}s_x^{(2)}},\ \left(s_y^{(2)}\right)^2=\frac{1}{\left(n^{(2)}-1\right)}\sum_{i=1}^{n^{(2)}}\left(y_i-\bar{y}^{(2)}\right)^2,\ \left(s_x^{(2)}\right)^2=\frac{1}{\left(n^{(2)}-1\right)}\sum_{i=1}^{n^{(2)}}\left(x_i-\bar{x}^{(2)}\right)^2,$$

$$\left(S_{yx}^{(2)}\right)=\frac{1}{\left(n^{(2)}-1\right)}\sum_{i=1}^{n^{(2)}}\left(y_i-\bar{y}^{(2)}\right)\left(x_i-\bar{x}^{(2)}\right)$$

เมื่อ f_1 คือ เศษส่วนการเลือกตัวอย่างในเฟสที่ 1

f_2 คือ เศษส่วนการเลือกตัวอย่างในเฟสที่ 2

$c_y^{(2)}$ คือ สัมประสิทธิ์ความแปรผันของตัวแปร Y ในตัวอย่างเฟสที่ 2

$c_x^{(2)}$ คือ สัมประสิทธิ์ความแปรผันของตัวแปร X ในตัวอย่างเฟสที่ 2

$r_{yx}^{(2)}$ คือ สัมประสิทธิ์สหสัมพันธ์ของตัวแปร X และ Y ในตัวอย่างเฟสที่ 2

4.1.2 ตัวประมาณผลคูณที่เสนอโดย Robson (1957)

เมื่อพิจารณาตัวประมาณผลคูณไม่เอนเอียงของ Robson (1957) ภายใต้การเลือกตัวอย่างแบบสองเฟสแล้ว จะได้

$$\bar{y}_{Rob}^{(2)}=\frac{\bar{y}^{(2)}\bar{x}^{(2)}}{\bar{x}^{(1)}}-f_2\frac{s_{yx}^{(2)}}{\bar{x}^{(1)}}$$

โดยมีความแปรปรวนดังนี้

พิจารณา $V\left(\bar{y}_{Rob}^{(2)}\right)\approx V_1\left\{E_2\left(\bar{y}_{Rob}^{(2)}\mid\mathbb{Z}\right)\right\}+E_1\left\{V_2\left(\bar{y}_{Rob}^{(2)}\mid\mathbb{Z}\right)\right\}$

จาก $\bar{y}_{Rob}^{(2)}$ เป็นตัวประมาณผลคูณไม่เอนเอียง ดังนั้น $E_2\left(\bar{y}_{Rob}^{(2)}\mid\mathbb{Z}\right)\rightarrow\bar{y}^{(1)}$

จะได้ $V_1\left\{E_2\left(\bar{y}_{Rob}^{(2)}\mid\mathbb{Z}\right)\right\}=V_1\left(\bar{y}^{(1)}\right)$

พิจารณา $E_1\left\{V_2\left(\bar{y}_{Rob}^{(2)}\mid\mathbb{Z}\right)\right\}$

โดยที่ $V_2\left(\bar{y}_{Rob}^{(2)}\mid\mathbb{Z}\right)=f_2\bar{Y}^2\left(C_Y^2+C_X^2+2\rho_{YX}C_YC_X\right)$

ดังนั้น

$$V\left(\bar{y}_{Rob}^{(2)}\right)\approx V_1\left(\bar{y}^{(1)}\right)+V_2\left(\bar{y}_{Rob}^{(2)}\right)$$
$$\approx\left(f_1\bar{Y}^2C_Y^2\right)+\left(f_2\bar{Y}^2\left(C_Y^2+C_X^2+2\rho_{YX}C_YC_X\right)\right)$$
$$V\left(\bar{y}_{Rob}^{(2)}\right)\approx\bar{Y}^2\left(f_1C_Y^2+f_2\left(C_Y^2+C_X^2+2\rho_{YX}C_YC_X\right)\right)$$

ตัวประมาณของ $V\left(\bar{y}_{Rob}^{(2)}\right)$ คือ

$$\hat{V}\left(\bar{y}_{Rob}^{(2)}\right)\approx\left(\bar{y}^{(2)}\right)^2\left(f_1\left(c_y^{(2)}\right)^2+f_2\left(\left(c_y^{(2)}\right)^2+\left(c_x^{(2)}\right)^2+2r_{yx}^{(2)}c_y^{(2)}c_x^{(2)}\right)\right)$$

4.1.3 ตัวประมาณที่เสนอโดย Mahanty & Mishra (2020)

เมื่อพิจารณาตัวประมาณไม่เอนเอียงของ Mahanty & Mishra (2020) ภายใต้การเลือกตัวอย่างแบบสองเฟสแล้ว จะได้

$$\bar{y}_{MU}^{(2)} = \lambda_0^{(2)}\left(\bar{y}^{(2)}\frac{\bar{x}^{(2)}}{\bar{x}^{(1)}} - f_2\frac{s_{yx}^{(2)}}{\bar{x}^{(1)}}\right) + \lambda_1^{(2)}\bar{x}^{(2)}$$

โดยมีความแปรปรวนดังนี้

พิจารณา $V\left(\bar{y}_{MU}^{(2)}\right) \approx V_1\left\{E_2\left(\bar{y}_{MU}^{(2)} \mid \mathbb{Z}\right)\right\} + E_1\left\{V_2\left(\bar{y}_{MU}^{(2)} \mid \mathbb{Z}\right)\right\}$

จาก $V_1\left\{E_2\left(\bar{y}_{MU}^{(2)} \mid \mathbb{Z}\right)\right\}$ เป็นตัวประมาณไม่เอนเอียง ดังนั้น $E_2\left(\bar{y}_{MU}^{(2)} \mid \mathbb{Z}\right) \to \bar{y}^{(1)}$

จะได้ $\quad V_1\left\{E_2\left(\bar{y}_{MU}^{(2)} \mid \mathbb{Z}\right)\right\} = V_1\left(\bar{y}^{(1)}\right)$

พิจารณา $\quad E_1\left\{V_2\left(\bar{y}_{MU}^{(2)} \mid \mathbb{Z}\right)\right\}$

โดยที่ $\quad V_2\left(\bar{y}_{MU}^{(2)} \mid \mathbb{Z}\right) = f_2\bar{Y}^2 C_X^2\left(1-\rho_{YX}^2\right)$

ดังนั้น

$$\begin{aligned} V\left(\bar{y}_{MU}^{(2)}\right) &\approx V_1\left(\bar{y}^{(1)}\right) + V_2\left(\bar{y}_{MU}^{(2)}\right) \\ &\approx \left(f_1\bar{Y}^2 C_Y^2\right) + \left(f_2\bar{Y}^2 C_X^2\left(1-\rho_{YX}^2\right)\right) \end{aligned}$$

$$V\left(\bar{y}_{MU}^{(2)}\right) \approx \bar{Y}^2\left(f_1 C_Y^2 + f_2 C_X^2\left(1-\rho_{YX}^2\right)\right)$$

ตัวประมาณของ $V\left(\bar{y}_{MU}^{(2)}\right)$ คือ

$$\hat{V}\left(\bar{y}_{MU}^{(2)}\right) \approx \left(\bar{y}^{(2)}\right)^2\left(f_1\left(C_y^{(2)}\right)^2 + f_2\left(C_x^{(2)}\right)^2\left(1-\left(r_{yx}^{(2)}\right)^2\right)\right)$$

เมื่อ $\quad \lambda_0 = \dfrac{\hat{R}^2\hat{V}\left(\bar{x}^{(2)}\right) - \hat{R}Cov\left(\bar{x}^{(2)}, \bar{y}_{Rob}^{(2)}\right)}{\hat{V}\left(\bar{y}_{Rob}^{(2)}\right) + R^2\hat{V}\left(\bar{x}^{(2)}\right) - 2\hat{R}Cov\left(\bar{x}^{(2)}, \bar{y}_{Rob}^{(2)}\right)}$

และ $\quad \lambda_1 = \hat{R}\cdot\dfrac{\hat{V}\left(\bar{y}_{Rob}^{(2)}\right) - \hat{R}Cov\left(\bar{x}^{(2)}, \bar{y}_{Rob}^{(2)}\right)}{\hat{V}\left(\bar{y}_{Rob}^{(2)}\right) + \hat{R}^2\hat{V}\left(\bar{x}^{(2)}\right) - 2\hat{R}Cov\left(\bar{x}^{(2)}, \bar{y}_{Rob}^{(2)}\right)}$

เมื่อ $\quad \hat{V}\left(\bar{x}^{(2)}\right) = f_2\left(\bar{x}^{(2)}\right)^2\left(c_x^{(2)}\right),\ Cov\left(\bar{x}^{(2)}, \bar{y}_{Rob}^{(2)}\right) = f_2\bar{y}^{(2)}\bar{x}^{(2)}\left(c_{yx} + \left(c_x^{(2)}\right)^2\right),$

$\quad \hat{R} = \dfrac{\bar{y}^{(2)}}{\bar{x}^{(2)}}$

4.2 เปรียบเทียบประสิทธิภาพของตัวประมาณ (Comparison of Efficiency)

เปรียบเทียบประสิทธิภาพของตัวประมาณโดยพิจารณาจากค่าประมาณความแปรปรวนของตัวประมาณ $\hat{V}(\bar{y}_1)-\hat{V}(\bar{y}_2)$

โดย $\hat{V}(\bar{y}_1)-\hat{V}(\bar{y}_2)>0$ แสดงว่าตัวประมาณค่าเฉลี่ยประชากร $\bar{y}_2$ มีประสิทธิภาพดีกว่า $\bar{y}_1$

$\hat{V}(\bar{y}_1)-\hat{V}(\bar{y}_2)<0$ แสดงว่าตัวประมาณค่าเฉลี่ยประชากร $\bar{y}_1$ มีประสิทธิภาพดีกว่า $\bar{y}_2$

4.2.1 การเปรียบเทียบ $\bar{y}_{MU}^{(2)}$ กับตัวประมาณค่าเฉลี่ยต่อหน่วย $\bar{y}_{Tin}^{(2)}$

$$
\begin{aligned}
V\left(\bar{y}_{Tin}^{(2)}\right)-V\left(\bar{y}_{MU}^{(2)}\right) &= \bar{Y}^2\left[f_1C_Y^2+f_2\left(C_Y^2+C_X^2-2\rho_{YX}C_YC_X\right)\right] \\
&\quad -\bar{Y}^2\left[f_1C_Y^2+f_2C_X^2\left(1-\rho_{YX}^2\right)\right] \\
&= \bar{Y}^2f_2\left[C_Y^2-\rho_{YX}C_X\left(2C_Y-\rho_{YX}C_X\right)\right]
\end{aligned}
$$

ดังนั้น ตัวประมาณ $\bar{y}_{MU}^{(2)}$ จะมีประสิทธิภาพสูงกว่าตัวประมาณค่าเฉลี่ยต่อหน่วย $\bar{y}_{Tin}^{(2)}$ ก็ต่อเมื่อ $\bar{Y}^2f_2\left[C_Y^2-\rho_{YX}C_X\left(2C_Y-\rho_{YX}C_X\right)\right]>0$

4.2.2 การเปรียบเทียบ $\bar{y}_{MU}^{(2)}$ กับตัวประมาณค่าเฉลี่ยต่อหน่วย $\bar{y}_{Rob}^{(2)}$

$$
\begin{aligned}
V\left(\bar{y}_{Rob}^{(2)}\right)-V\left(\bar{y}_{MU}^{(2)}\right) &= \bar{Y}^2\left[f_1C_Y^2+f_2\left(C_Y^2+C_X^2+2\rho_{YX}C_YC_X\right)\right] \\
&\quad -\bar{Y}^2\left[f_1C_Y^2+f_2C_X^2\left(1-\rho_{YX}^2\right)\right] \\
&= \bar{Y}^2f_2\left[C_Y^2+\rho_{YX}C_X\left(2C_Y+\rho_{YX}C_X\right)\right]
\end{aligned}
$$

ดังนั้น ตัวประมาณ $\bar{y}_{MU}^{(2)}$ จะมีประสิทธิภาพสูงกว่าตัวประมาณค่าเฉลี่ยต่อหน่วย $\bar{y}_{Rob}^{(2)}$ ก็ต่อเมื่อ $\bar{Y}^2f_2\left[C_Y^2+\rho_{YX}C_X\left(2C_Y+\rho_{YX}C_X\right)\right]>0$

4.3 การอธิบายเชิงตัวเลข (Numerically Illustrations)

ในการจำลองข้อมูลของแต่ละตัวประมาณ จะดำเนินการทำซ้ำทั้งหมด 5,000 รอบ

คำนวณค่า $\bar{y}_i=\dfrac{\sum_{k=1}^{5,000}\bar{y}_k}{5,000}$ เมื่อ i คือ ตัวประมาณที่ศึกษาตัวที่ i

และ $$Bi\hat{a}s(\bar{y}_i)=\frac{\sum_{k=1}^{5,000}Bi\hat{a}s(\bar{y}_k)}{5,000}$$

และ $$\hat{V}(\bar{y}_i)=\frac{\sum_{k=1}^{5000}\hat{V}(\bar{y}_k)}{5,000}$$

4.3.1 ผลการวิเคราะห์ข้อมูล

ตารางที่ 1 คุณลักษณะของประชากร (X และ Y มีความสัมพันธ์เชิงเส้นตรง ในทิศทางเดียวกัน)

$N = 1,000$			
	$\rho_{YX} = 0.8$	$\rho_{YX} = 0.6$	$\rho_{YX} = 0.3$
$\bar{X}$	500.1441	499.5214	500.2281
$\bar{Y}$	100.5245	98.6965	101.3629
S_X^2	99.8032	100.7538	93.5135
S_Y^2	397.2334	406.2266	364.3126
S_{YX}	158.3945	120.2792	53.5595
C_X^2	0.0003989	0.0004038	0.00037371
C_Y^2	0.0393099	0.0417027	0.0355
ρ_{YX}	0.7955	0.5945	0.2902

ตารางที่ 2 คุณลักษณะของประชากร (X และ Y มีความสัมพันธ์เชิงเส้นตรง ในทิศทางตรงกันข้าม)

$N = 1,000$			
	$\rho_{YX} = -0.8$	$\rho_{YX} = -0.6$	$\rho_{YX} = -0.3$
$\bar{X}$	499.7245	500.3034	499.8307
$\bar{Y}$	100.5245	98.6965	100.4322
S_X^2	100.6014	102.6603	101.6708
S_Y^2	397.2334	406.2266	378.0945
S_{YX}	-159.3922	-123.4567	-58.49173
C_X^2	0.0004028	0.0004101	0.0004069
C_Y^2	0.0393099	0.0417027	0.0374847
ρ_{YX}	-0.7973	-0.6045	-0.2983

ตารางที่ 3 ค่าเฉลี่ยและค่าความเอนเอียงของตัวประมาณอัตราส่วน ภายใต้การเลือกตัวอย่างสองเฟส โดยที่ประชากรมีความสัมพันธ์ในทิศทางเดียวกัน

$n^{(1)}$	$n^{(2)}$	%	$\bar{y}_R^{(2)}$	$\bar{y}_{Tin}^{(2)}$	$\bar{y}_{MU}^{(2)}$	$Bias\left(\bar{y}_R^{(2)}\right)$	$Bias\left(\bar{y}_{Tin}^{(2)}\right)$	$Bias\left(\bar{y}_{MU}^{(2)}\right)$
$\rho_{YX} = 0.8$								
50	10	20	101.3917	101.4115	101.4201	-0.0288	0.0486	0.0573
100	20	20	101.3365	101.3466	101.3468	-0.0263	-0.0162	-0.0153
100	50	50	101.3139	101.3165	101.3154	-0.0489	-0.0467	-0.0475
200	50	25	101.3691	101.3729	101.3741	-0.0102	0.0100	0.0112
200	100	50	101.3755	101.3768	101.3768	-0.0126	0.0139	0.0139
500	100	20	101.3027	101.3048	101.3021	-0.0601	-0.0581	-0.0607
500	200	40	101.3648	101.3655	101.3641	-0.0019	0.0026	0.0012
500	300	60	101.3633	101.3636	101.3627	-0.0004	0.0007	-0.0002
$\rho_{YX} = 0.6$								
50	10	20	101.3928	101.4071	101.4063	-0.0299	0.0442	0.0434
100	20	20	101.3686	101.3758	101.3787	-0.0157	0.0129	0.0158
100	50	50	101.3549	101.3567	101.3612	-0.0079	-0.0061	-0.0017
200	50	25	101.4082	101.4108	101.4110	-0.0453	0.0479	0.0481
200	100	50	101.3492	101.3501	101.3493	-0.0136	-0.0127	-0.0135
500	100	20	101.3676	101.3690	101.3660	-0.0047	0.0062	0.0031
500	200	40	101.3617	101.3622	101.3615	-0.0011	-0.0006	-0.0013
500	300	60	101.3713	101.3716	101.3723	-0.0084	0.0087	0.0094
$\rho_{YX} = 0.3$								
50	10	20	101.3000	101.3058	101.2993	-0.0571	-0.0636	-0.0635
100	20	20	101.2686	101.2714	101.2705	-0.0942	-0.0914	-0.0923
100	50	50	101.3808	101.3815	101.3815	-0.0179	0.0186	0.0186
200	50	25	101.3596	101.3607	101.3624	-0.0032	-0.0021	-0.0004
200	100	50	101.3629	101.3633	101.3647	-0.0001	0.0004	0.0018
500	100	20	101.3556	101.3561	101.3560	-0.0073	-0.0067	-0.0068
500	200	40	101.3526	101.3529	101.3517	-0.0102	-0.0100	-0.0111
500	300	60	101.3531	101.3532	101.3519	-0.0097	-0.0096	-0.0109

** % คือ สัดส่วนการเลือกตัวอย่างเฟสที่2 จากเฟสที่1*

จากตารางที่ 3 ภายใต้การเลือกตัวอย่างสองเฟส โดยที่ประชากรมีความสัมพันธ์ในทิศทางเดียวกัน พบว่า ค่าความเอนเอียงของทั้งสามตัวประมาณ และทุกขนาดตัวอย่างมีค่าเข้าใกล้ศูนย์ ซึ่งหมายความว่า ตัวประมาณที่ผู้วิจัยพัฒนาเป็นตัวประมาณไม่เอนเอียงภายใต้การเลือกตัวอย่างสองเฟส

ตารางที่ 4 ค่าเฉลี่ยและค่าความเอนเอียงของตัวประมาณผลคูณ ภายใต้การเลือกตัวอย่างสองเฟส โดยที่ประชากรมีความสัมพันธ์ในทิศทางตรงกันข้าม

$n^{(1)}$	$n^{(2)}$	%	$\bar{y}_P^{(2)}$	$\bar{y}_{Rob}^{(2)}$	$\bar{y}_{MU}^{(2)}$	$Bias\left(\bar{y}_P^{(2)}\right)$	$Bias\left(\bar{y}_{Rob}^{(2)}\right)$	$Bias\left(\bar{y}_{MU}^{(2)}\right)$
				$\rho_{YX} = -0.8$				
50	10	20	101.4407	101.4641	101.4689	-0.0778	0.1013	0.1061
100	20	20	101.4471	101.4589	101.4639	-0.0842	0.0960	0.1010
100	50	50	101.3568	101.3597	101.3611	-0.0061	-0.0031	-0.0018
200	50	25	101.3423	101.3467	101.3480	-0.0205	-0.0161	-0.0148
200	100	50	101.3307	101.3322	101.3328	-0.0321	-0.0306	-0.0300
500	100	20	101.3384	101.3408	101.3391	-0.0244	-0.0221	-0.0237
500	200	40	101.3675	101.3684	101.3664	-0.0046	0.0055	0.0035
500	300	60	101.3608	101.3612	101.3622	-0.0020	-0.0016	-0.0006
				$\rho_{YX} = -0.6$				
50	10	20	101.4391	101.4641	101.4689	-0.0762	0.1013	0.1061
100	20	20	101.3815	101.3903	101.3926	-0.0186	0.0273	0.0297
100	50	50	101.3163	101.3185	101.3206	(-0.0465)	-0.0443	-0.0423
200	50	25	101.4338	101.4371	101.4482	-0.0709	0.0742	0.0853
200	100	50	101.3727	101.3738	101.3747	-0.0098	0.0109	0.0118
500	100	20	101.376	101.3778	101.3806	-0.0131	0.0149	0.0177
500	200	40	101.3625	101.3631	101.3610	(-0.0003)	0.0002	-0.0018
500	300	60	101.3581	101.3584	101.3596	(-0.0047)	-0.0044	0.0032
				$\rho_{YX} = -0.3$				
50	10	20	101.1522	101.1610	101.1561	(-0.2106)	-0.2018	-0.2067
100	20	20	101.3175	101.3219	101.3228	(-0.0453)	-0.0409	-0.0401
100	50	50	101.4184	101.4195	101.4170	-0.0555	0.0566	0.0541
200	50	25	101.3901	101.3918	101.3932	-0.0272	0.0289	0.0303
200	100	50	101.3139	101.3144	101.3126	(-0.0489)	-0.0484	-0.0502
500	100	20	101.3584	101.3593	101.3580	(-0.0044)	-0.0035	-0.0048
500	200	40	101.3384	101.3388	101.3403	(-0.0244)	-0.0240	-0.0225
500	300	60	101.3597	101.3599	101.3593	(-0.0031)	-0.0029	-0.0035

** % คือสัดส่วนการเลือกตัวอย่างเฟสที่2 จากเฟสที่1*

จากตารางที่ 4 ภายใต้การเลือกตัวอย่างสองเฟส โดยที่ประชากรมีความสัมพันธ์ในทิศทางตรงกันข้าม พบว่า ค่าความเอนเอียงของทั้งสามตัวประมาณ และทุกขนาดตัวอย่างมีค่าเข้าใกล้ศูนย์ ซึ่งหมายความว่า ตัวประมาณที่ผู้วิจัยพัฒนาเป็นตัวประมาณไม่เอนเอียงภายใต้การเลือกตัวอย่างสองเฟส

ตารางที่ 5 ค่าความแปรปรวนและค่าประมาณความแปรปรวนของตัวประมาณอัตราส่วน ภายใต้การเลือกตัวอย่างสองเฟส โดยที่ประชากรมีความสัมพันธ์กันในทิศทางเดียวกัน

$\rho_{YX}=0.8$							
$n^{(1)}$	$n^{(2)}$	$V\left(\bar{y}_R^{(2)}\right)$	$V\left(\bar{y}_{Tin}^{(2)}\right)$	$V\left(\bar{y}_{MU}^{(2)}\right)$	$\hat{V}\left(\bar{y}_R^{(2)}\right)$	$\hat{V}\left(\bar{y}_{Tin}^{(2)}\right)$	$\hat{V}\left(\bar{y}_{MU}^{(2)}\right)$
50	10	31.6682	31.6682	7.0331	31.1040	31.1040	6.8920
100	20	15.6519	15.6519	3.3343	15.6806	15.6806	3.3369
100	50	6.3721	6.3721	3.2927	6.4107	6.4107	3.3123
200	50	6.0971	6.0971	1.4780	6.0901	6.0901	1.4759
200	100	3.0038	3.0038	1.4641	3.0016	3.0016	1.4631
500	100	2.8389	2.8389	0.3754	2.8438	2.8438	0.3759
500	200	1.2922	1.2922	0.3684	1.2910	1.2910	0.3680
500	300	0.7767	0.7767	0.3661	0.7767	0.7767	0.3661
$\rho_{YX}=0.6$							
$n^{(1)}$	$n^{(2)}$	$V\left(\bar{y}_R^{(2)}\right)$	$V\left(\bar{y}_{Tin}^{(2)}\right)$	$V\left(\bar{y}_{MU}^{(2)}\right)$	$\hat{V}\left(\bar{y}_R^{(2)}\right)$	$\hat{V}\left(\bar{y}_{Tin}^{(2)}\right)$	$\hat{V}\left(\bar{y}_{MU}^{(2)}\right)$
50	10	32.8579	32.8579	7.1196	32.8746	32.8746	7.0980
100	20	16.2468	16.2468	3.3776	16.3244	16.3244	3.3870
100	50	6.5208	6.5208	3.3035	6.5467	6.5467	3.3158
200	50	6.3202	6.3202	1.4943	6.3172	6.3172	1.4930
200	100	3.0782	3.0782	1.4696	3.0703	3.0703	1.4658
500	100	2.9579	2.9579	0.3841	2.9562	2.9562	0.3837
500	200	1.3369	1.3369	0.3717	1.3379	1.3379	0.3719
500	300	0.7965	0.7965	0.3676	0.7953	0.7953	0.3670
$\rho_{YX}=0.3$							
$n^{(1)}$	$n^{(2)}$	$V\left(\bar{y}_R^{(2)}\right)$	$V\left(\bar{y}_{Tin}^{(2)}\right)$	$V\left(\bar{y}_{MU}^{(2)}\right)$	$\hat{V}\left(\bar{y}_R^{(2)}\right)$	$\hat{V}\left(\bar{y}_{Tin}^{(2)}\right)$	$\hat{V}\left(\bar{y}_{MU}^{(2)}\right)$
50	10	34.6376	34.6376	7.2032	34.8735	34.8735	7.2239
100	20	17.1366	17.1366	3.4194	17.3747	17.3747	3.4583
100	50	6.7433	6.7433	3.3140	6.7472	6.7472	3.3164
200	50	6.6539	6.6539	1.5099	6.6586	6.6586	1.5103
200	100	3.1894	3.1894	1.4748	3.1901	3.1901	1.4748
500	100	3.1359	3.1359	0.3924	3.1378	3.1378	0.3923
500	200	1.4036	1.4036	0.3748	1.4007	1.4007	0.3740
500	300	0.8262	0.8262	0.3690	0.8263	0.8263	0.3690

จากตารางที่ 5 พบว่า เมื่อระดับความสัมพันธ์ของตัวแปรที่สนใจศึกษา Y และตัวแปรช่วย X มีค่าสูงขึ้น และขนาดตัวอย่างมีขนาดเพิ่มขึ้น จะเห็นได้ว่า ค่าความแปรปรวนของตัวประมาณมีค่าลดลง หากพิจารณาค่าความแปรปรวนในขนาดตัวอย่างเดียวกัน เมื่อระดับความสัมพันธ์ของตัวแปรที่สนใจศึกษา Y และตัวแปรช่วย X เพิ่มขึ้น จะเห็นได้ว่าค่าความแปรปรวนของตัวประมาณจะลดลงเช่นเดียวกัน

เมื่อพิจารณาเปรียบเทียบประสิทธิภาพของตัวประมาณ พบว่า $\hat{V}\left(\bar{y}_{Tin}^{(2)}\right)-\hat{V}\left(\bar{y}_{MU}^{(2)}\right)>0$ และ $\hat{V}\left(\bar{y}_{R}^{(2)}\right)-\hat{V}\left(\bar{y}_{MU}^{(2)}\right)>0$ จึงสามารถสรุปได้ว่า $\bar{y}_{MU}^{(2)}$ ตัวประมาณที่ผู้วิจัยนำเสนอมีประสิทธิภาพมากกว่า $\bar{y}_{Tin}^{(2)}$ และ $\bar{y}_{R}^{(2)}$

ตารางที่ 6 ค่าความแปรปรวนและค่าประมาณความแปรปรวนของตัวประมาณผลคูณ ภายใต้การเลือกตัวอย่างสองเฟส โดยที่ประชากรมีความสัมพันธ์กันในทิศทางตรงกันข้าม

$n^{(1)}$	$n^{(2)}$	$V\left(\bar{y}_P^{(2)}\right)$	$V\left(\bar{y}_{Rob}^{(2)}\right)$	$V\left(\bar{y}_{MU}^{(2)}\right)$	$\hat{V}\left(\bar{y}_P^{(2)}\right)$	$\hat{V}\left(\bar{y}_{Rob}^{(2)}\right)$	$\hat{V}\left(\bar{y}_{MU}^{(2)}\right)$
				$\rho_{YX} = -0.8$			
50	10	31.6184	31.6184	7.0335	31.6294	31.6294	7.0236
100	20	15.6270	15.6270	3.3346	15.6307	15.6307	3.3339
100	50	6.3659	6.3659	3.2928	6.3973	6.3973	3.3087
200	50	6.0878	6.0878	1.4781	6.0650	6.0650	1.4727
200	100	3.0007	3.0007	1.4642	2.9988	2.9988	1.4631
500	100	2.8339	2.8339	0.3754	2.8353	2.8353	0.3755
500	200	1.2904	1.2904	0.3684	1.2899	1.2899	0.3683
500	300	0.7759	0.7759	0.3661	0.7747	0.7747	0.3656
				$\rho_{YX} = -0.6$			
50	10	32.7976	32.7976	7.1202	33.2543	33.2543	7.1928
100	20	16.2166	16.2166	3.3779	16.1911	16.1911	3.3643
100	50	6.5132	6.5132	3.3036	6.5354	6.5354	3.3147
200	50	6.3089	6.3089	1.4944	6.3053	6.3053	1.4931
200	100	3.0744	3.0744	1.4696	3.0723	3.0723	1.4687
500	100	2.9518	2.9518	0.3841	2.9506	2.9506	0.3838
500	200	1.3346	1.3346	0.3717	1.3323	1.3323	0.3710
500	300	0.7955	0.7955	0.3676	0.7960	0.7960	0.3678
					$\rho_{YX} = -0.3$		
50	10	34.5683	34.5683	7.2037	34.5562	34.5562	7.1673
100	20	17.1020	17.1020	3.4196	17.0621	17.0621	3.4044
100	50	6.7346	6.7346	3.3140	6.7322	6.7322	3.3126
200	50	6.6409	6.6409	1.5100	6.6245	6.6245	1.5059
200	100	3.1851	3.1851	1.4748	3.1842	3.1842	1.4743
500	100	3.1289	3.1289	0.3924	3.1359	3.1359	0.3928
500	200	1.4010	1.4010	0.3748	1.4018	1.4018	0.3750
500	300	0.8250	0.8250	0.3690	0.8247	0.8247	0.3688

จากตารางที่ 6 พบว่า เมื่อระดับความสัมพันธ์ของตัวแปรที่สนใจศึกษา Y และตัวแปรช่วย X มีค่าสูงขึ้น และขนาดตัวอย่างมีขนาดเพิ่มขึ้น จะเห็นได้ว่า ค่าความแปรปรวนของตัวประมาณมีค่าลดลง หากพิจารณาค่าความแปรปรวนในขนาดตัวอย่างเดียวกัน เมื่อระดับความสัมพันธ์ของตัวแปรที่สนใจศึกษา Y และตัวแปรช่วย X เพิ่มขึ้น จะเห็นได้ว่าค่าความแปรปรวนของตัวประมาณจะลดลงเช่นเดียวกัน

เมื่อพิจารณาเปรียบเทียบประสิทธิภาพของตัวประมาณพบว่า $\hat{V}\left(\bar{y}_{Rob}^{(2)}\right)-\hat{V}\left(\bar{y}_{MU}^{(2)}\right)>0$ และ $\hat{V}\left(\bar{y}_{P}^{(2)}\right)-\hat{V}\left(\bar{y}_{MU}^{(2)}\right)>0$ จึงสามารถสรุปได้ว่า $\bar{y}_{MU}^{(2)}$ ตัวประมาณที่ผู้วิจัยนำเสนอมีประสิทธิภาพมากกว่า $\bar{y}_{Rob}^{(2)}$ และ $\bar{y}_{P}^{(2)}$

ผู้วิจัยเปรียบเทียบประสิทธิภาพของตัวประมาณในเชิงทฤษฎีผ่านการการอธิบายเชิงตัวเลขดังนี้

ตารางที่ 7 เปรียบเทียบประสิทธิภาพของตัวประมาณในเชิงทฤษฎีผ่านการอธิบายเชิงตัวเลขของตัวประมาณ ภายใต้การเลือกตัวอย่างสองเฟส

Efficiency Comparisons								
$n^{(1)}$	$n^{(2)}$	$\hat{V}(\bar{y}_1)-\hat{V}(\bar{y}_2)$	$\rho_{YX}=0.8$	$\rho_{YX}=0.6$	$\rho_{YX}=0.3$	$\rho_{YX}=-0.8$	$\rho_{YX}=-0.6$	$\rho_{YX}=-0.3$
50	10	$\hat{V}(\bar{y}_{Tin}^{(2)})-\hat{V}(\bar{y}_{MU}^{(2)})$	24.6549	25.5443	27.2677	*	*	*
		$\hat{V}(\bar{y}_{Rob}^{(2)})-\hat{V}(\bar{y}_{MU}^{(2)})$	*	*	*	24.6112	25.5535	27.4772
100	20	$\hat{V}(\bar{y}_{Tin}^{(2)})-\hat{V}(\bar{y}_{MU}^{(2)})$	12.2381	12.8821	13.8350	*	*	*
		$\hat{V}(\bar{y}_{Rob}^{(2)})-\hat{V}(\bar{y}_{MU}^{(2)})$	*	*	*	12.3443	12.8778	13.7319
100	50	$\hat{V}(\bar{y}_{Tin}^{(2)})-\hat{V}(\bar{y}_{MU}^{(2)})$	3.0918	3.2289	3.4260	*	*	*
		$\hat{V}(\bar{y}_{Rob}^{(2)})-\hat{V}(\bar{y}_{MU}^{(2)})$	*	*	*	3.0760	3.2115	3.42799
200	50	$\hat{V}(\bar{y}_{Tin}^{(2)})-\hat{V}(\bar{y}_{MU}^{(2)})$	4.6267	4.8060	5.1138	*	*	*
		$\hat{V}(\bar{y}_{Rob}^{(2)})-\hat{V}(\bar{y}_{MU}^{(2)})$	*	*	*	4.5899	4.8192	5.1050
200	100	$\hat{V}(\bar{y}_{Tin}^{(2)})-\hat{V}(\bar{y}_{MU}^{(2)})$	1.5398	1.6087	1.7122	*	*	*
		$\hat{V}(\bar{y}_{Rob}^{(2)})-\hat{V}(\bar{y}_{MU}^{(2)})$	*	*	*	1.5363	1.6027	1.7073
500	100	$\hat{V}(\bar{y}_{Tin}^{(2)})-\hat{V}(\bar{y}_{MU}^{(2)})$	2.4605	2.5757	2.7478	*	*	*
		$\hat{V}(\bar{y}_{Rob}^{(2)})-\hat{V}(\bar{y}_{MU}^{(2)})$	*	*	*	2.4624	2.5666	2.7421
500	200	$\hat{V}(\bar{y}_{Tin}^{(2)})-\hat{V}(\bar{y}_{MU}^{(2)})$	0.9247	0.9660	1.0308	*	*	*
		$\hat{V}(\bar{y}_{Rob}^{(2)})-\hat{V}(\bar{y}_{MU}^{(2)})$	*	*	*	0.9215	0.9663	1.0249
500	300	$\hat{V}(\bar{y}_{Tin}^{(2)})-\hat{V}(\bar{y}_{MU}^{(2)})$	0.4105	0.4282	0.4575	*	*	*
		$\hat{V}(\bar{y}_{Rob}^{(2)})-\hat{V}(\bar{y}_{MU}^{(2)})$	*	*	*	0.4101	0.4277	0.4557

** เป็นตัวประมาณที่ประชากรมีความสัมพันธ์แตกต่างกัน จึงไม่ทำการเปรียบเทียบ*

จากตารางที่ 7 เมื่อเปรียบเทียบประสิทธิภาพของตัวประมาณในเชิงทฤษฎีผ่านการการอธิบายเชิงตัวเลขของตัวประมาณ ภายใต้การเลือกตัวอย่างสองเฟส พบว่า ตัวประมาณที่ผู้วิจัยนำเสนอ $\bar{y}_{MU}^{(2)}$ มีประสิทธิภาพสูงกว่าตัวประมาณอื่นที่ศึกษา ทั้งในประชากรมีความสัมพันธ์กันในทิศทางเดียวกันและตรงกันข้าม ในทุกขนาดตัวอย่าง

บทที่ 5
สรุป อภิปรายผล และข้อเสนอแนะ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อศึกษาและพัฒนาตัวประมาณไม่เอนเอียงโดยการใช้สารสนเทศของตัวแปรช่วยสำหรับค่าเฉลี่ยประชากรจำกัด ภายใต้การเลือกตัวอย่างสองเฟส และเปรียบเทียบประสิทธิภาพของตัวประมาณที่ศึกษากับตัวประมาณอัตราส่วนและตัวประมาณผลคูณแบบทั่วไปโดยใช้ความแปรปรวน (Variance) เป็นเกณฑ์

5.1 สรุปผลการวิจัย

ในการวิจัยครั้งนี้ผู้วิจัยพัฒนาตัวประมาณอัตราส่วนไม่เอนเอียงและตัวประมาณผลคูณไม่เอนเอียง ที่เสนอโดย Tin (1965) Robson (1957) และ Mahanty และ Mishra (2020) ภายใต้การเลือกตัวอย่างแบบสองเฟส ดังนี้

ตัวประมาณอัตราส่วนที่เสนอโดย Tin (1965)

$$\bar{y}_{Tin}^{(2)} = \bar{y}^{(2)}\frac{\bar{x}^{(1)}}{\bar{x}^{(2)}}\left[1+f_2\left(\frac{s_{yx}^{(2)}}{\bar{x}^{(2)}\bar{y}^{(2)}}-\frac{\left(s_x^{(2)}\right)^2}{\left(\bar{x}^{(2)}\right)^2}\right)\right]$$

โดยมีค่าความแปรปรวนดังนี้

$$V\left(\bar{y}_{Tin}^{(2)}\right)\approx\bar{Y}^2\left[f_1C_Y^2+f_2\left(C_Y^2+C_X^2-2\rho_{YX}C_YC_X\right)\right]$$

ตัวประมาณผลคูณที่เสนอโดย Robson (1957)

$$\bar{y}_{Rob}^{(2)}=\frac{\bar{y}^{(2)}\bar{x}^{(2)}}{\bar{x}^{(1)}}-f_2\frac{s_{yx}^{(2)}}{\bar{x}^{(1)}}$$

โดยมีค่าความแปรปรวนดังนี้

$$V\left(\bar{y}_{Rob}^{(2)}\right)\approx\bar{Y}^2\left[f_1C_Y^2+f_2\left(C_Y^2+C_X^2+2\rho_{YX}C_YC_X\right)\right]$$

ตัวประมาณที่เสนอโดย Mahanty & Mishra (2020)

$$\bar{y}_{MU}^{(2)}=\lambda_0^{(2)}\left(\bar{y}^{(2)}\frac{\bar{x}^{(2)}}{\bar{x}^{(1)}}-f_2\frac{s_{yx}^{(2)}}{\bar{x}^{(1)}}\right)+\lambda_1^{(2)}\bar{x}^{(2)}$$

โดยมีค่าความแปรปรวนดังนี้

$$V\left(\bar{y}_{MU}^{(2)}\right)_{opt}\approx\bar{Y}^2\left[f_1C_Y^2+f_2C_X^2\left(1-\rho_{YX}^2\right)\right]$$

ข้อมูลที่ผู้วิจัยได้นำมาใช้ในการวิเคราะห์เป็นการจำลองข้อมูลขึ้น โดยใช้โปรแกรมทางสถิติ (RStudio) สร้างประชากรขนาด $N=1,000$ หน่วย โดย $Y\sim N\left(500,100^2\right)$ และ $X\sim N\left(100,20^2\right)$ กำหนดค่าความสัมพันธ์ของประชากร 3 ระดับในกรณี Y และ X มีความสัมพันธ์กันทางบวกและทางลบ

1 ประชากรมีความสัมพันธ์กันสูง $\rho_{YX}=0.8$ และ $\rho_{YX}=-0.8$

2 ประชากรมีความสัมพันธ์กันปานกลาง $\rho_{YX}=0.6$ และ $\rho_{YX}=-0.6$

3 ประชากรมีความสัมพันธ์กันน้อย $\rho_{YX}=0.3$ และ $\rho_{YX}=-0.3$

จากประชากรที่มีความสัมพันธ์กันทั้ง 3 ระดับจะทำการสุ่มข้อมูลภายใต้การเลือกตัวอย่างสองเฟส 8 รูปแบบ ดังนี้

1. $n^{(1)}=50$ และ $n^{(2)}=10$
2. $n^{(1)}=100$ และ $n^{(2)}=20$
3. $n^{(1)}=100$ และ $n^{(2)}=50$
4. $n^{(1)}=200$ และ $n^{(2)}=50$
5. $n^{(1)}=200$ และ $n^{(2)}=100$
6. $n^{(1)}=500$ และ $n^{(2)}=100$
7. $n^{(1)}=500$ และ $n^{(2)}=200$
8. $n^{(1)}=500$ และ $n^{(2)}=300$

จากนั้นทำการเปรียบเทียบประสิทธิภาพของตัวประมาณที่ผู้วิจัยพัฒนา โดยใช้ค่าความแปรปรวนเป็นเกณฑ์

จากผลการเปรียบเทียบตัวประมาณอัตราส่วนไม่เอนเอียงและตัวประมาณผลคูณไม่เอียนเอียง พบว่า เมื่อระดับความสัมพันธ์ และขนาดตัวอย่างมีขนาดเพิ่มขึ้น จะเห็นว่าค่าความแปรปรวนของตัวประมาณมีค่าลดลง หากพิจารณาค่าความแปรปรวนในขนาดตัวอย่างเดียวกันแต่ระดับความสัมพันธ์ต่างกัน เมื่อระดับความสัมพันธ์เพิ่มขึ้นจะเห็นว่าค่าความแปรปรวนของตัวประมาณจะลดลงเช่นเดียวกัน

เมื่อพิจารณาเปรียบเทียบประสิทธิภาพของตัวประมาณพบว่า $\hat{V}\left(\bar{y}_{Tin}^{(2)}\right)-\hat{V}\left(\bar{y}_{MU}^{(2)}\right)>0$ และ $\hat{V}\left(\bar{y}_{Rob}^{(2)}\right)-\hat{V}\left(\bar{y}_{MU}^{(2)}\right)>0$ จึงสามารถสรุปได้ว่า $\bar{y}_{MU}^{(2)}$ มีประสิทธิภาพมากกว่า $\bar{y}_{Tin}^{(2)}$ และ $\bar{y}_{Rob}^{(2)}$

5.2 อภิปรายผล

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อศึกษาและพัฒนาตัวประมาณไม่เอนเอียงโดยการใช้สารสนเทศของตัวแปรช่วยสำหรับค่าเฉลี่ยประชากรจำกัด ภายใต้การเลือกตัวอย่างสองเฟส และเปรียบเทียบประสิทธิภาพของตัวประมาณที่ศึกษากับตัวประมาณอัตราส่วนและตัวประมาณผลคูณแบบทั่วไปโดยใช้ความแปรปรวน (Variance) เป็นเกณฑ์ พบว่า ตัวประมาณที่ผู้วิจัยพัฒนาที่เสนอโดย Mahanty & Mishra (2020) มีประสิทธิภาพสูงกว่าตัวประมาณอัตราส่วนไม่เอนเอียงที่เสนอโดย Tin (1965) และ ตัวประมาณผลคูณไม่เอนเอียงที่เสนอโดย Robson (1957) เนื่องจากมีค่าความแปรปรวนที่ต่ำที่สุดในทุก ๆ ระดับความสัมพันธ์ ทุกขนาดตัวอย่าง และตัวประมาณที่ผู้วิจัยพัฒนา ที่เสนอโดย Mahanty & Mishra (2020) ยังมีประสิทธิภาพสูงกว่าตัวประมาณอื่น ๆ ในประชากรมีความสัมพันธ์กันในทิศทางเดียวกันและทิศทางตรงกันข้าม ซึ่งสอดคล้องกับงานวิจัยของ B. Mahanty และ G. Mishra ได้ทำงานวิจัยเรื่อง ตัวประมาณที่ไม่เอนเอียงของค่าเฉลี่ยประชากรแบบจำกัดโดยใช้สารสนเทศของตัวแปรช่วย ผลการศึกษาพบว่าผู้วิจัยสร้างตัวประมาณที่ไม่เอนเอียงขึ้นโดยใช้ผลรวมเชิงเส้นของตัวประมาณตัวแปรที่ศึกษาและ

ค่าเฉลี่ยต่อหน่วยของตัวแปรช่วยภายใต้การสุ่มตัวอย่างอย่างง่ายแบบไม่คืนที่ เปรียบเทียบประสิทธิภาพของตัวประมาณภายใต้สภาวะที่เหมาะสมที่สุดกับตัวประมาณค่าเฉลี่ยต่อหน่วย ตัวประมาณอัตราส่วนไม่เอนเอียง ตัวประมาณผลคูณไม่เอนเอียง และตัวประมาณการถดถอย ทั้งในทางทฤษฎีและเชิงตัวเลข

5.3 ข้อเสนอแนะ

ในการวิจัยครั้งนี้ผู้วิจัยได้จำลองข้อมูลขึ้นมา โดยใช้ข้อมูลมีการแจกแจงแบบปกติเท่านั้น ในการวิจัยครั้งต่อไปควรจำลองข้อมูลให้มีการแจกแจงแบบอื่น ๆ เช่น ข้อมูลที่มีลักษณะเบ้ซ้าย หรือเบ้ขวา

บรรณานุกรม

จิราภรณ์ ธรรมสาโรช. (2560). ตัวประมาณอัตราส่วนร่วมกับผลคูณแบบ ปรับปรุงภายใต้การเลือกตัวอย่างสองเฟส. วารสารวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยอุบลราชธานี, *19*(1), 84-96.

ณภัทณ์จันทร์ ด่านสวัสดิ์. (2563). ตัวประมาณอัตราส่วนร่วมกับผลคูณแบบเลขชี้กำลัง สำหรับค่าเฉลี่ยประชากรในการเลือกตัวอย่างสุ่มแบบง่าย โดยใช้สารสนเทศจากตัวแปรช่วย. วารสารวิทยาศาสต์บูรพา, *25*(2), 563-577.

นิภาพร ชุติมันต์. (2557). วิธีการเลือกตัวอย่าง. หจก.อภิชาติการพิมพ์.

ภาควิชาคณิตศาสตร์. (2558). ระเบียบวิธีการทางสถิติทั่วไป. หจก.อภิชาติการพิมพ์.

วิชชญาพร สิงห์นวล, & จิราวัลย์ จิตรถเวช. (2559). การเปรียบเทียบประสิทธิภาพของตัวประมาณค่าต่างๆกับตัวประมาณค่าของซิลส์ที่ปรับปรุงใหม่. วารสารวิจัยและพัฒนา มจธ., *39*(3), 327-336.

Irfan, M., Javed, M., Bhatti, S. H., Raza, M. A., & Ahmad, T. (2020). Almost unbiased optimum estimators for population mean using dual auxiliary information. *Journal of King Saud University – Science*, *32*, 2835-2844.

Lohr, S. L. (2022). *Sampling: design and analysis* (Vol. 3). 6000 Broken Sound Parkway NWm Suite 300, Boca Raton, FL 33487-2742.

Mahanty, B., & Mishra, G. (2020). An Unbiased Estimator of Finite Population Mean Using Auxiliary. *Information Journal of Statistical Theory and Applications*, *19*(4), 534-539.

Robson, D. S. (1957). Applications of Multivariate Polykays to the Theory of Unbiased Ratio-Type Estimation. *Journal of the American Statistical Association*, *52*, 511-522.

Singh, H. P., & Pal, S. K. (2015). A New Chain Ratio-Ratio-Type Exponential Estimator Using Auxiliary Information in Sample Surveys. *International Journal of Mathematics And Its Applications*, *3*(4), 37-46.

Tin, M. (1965). Comparison of Some Ratio Estimators. *Journal of the American Statistical Association*, *60*, 294-307.

Vishwakarma, G. K., Singh, R., Gupta, P. C., & Pareek, S. (2016). Improved ratio and product type estimators of finite population mean in simple random sampling.

Revista Investigacion Operacional, *36*(3), 70-76.

Vishwakarma, G. K., & Zeeshan, S. M. (2021). Generalized Ratio-cum-Product Estimator for Finite Population Mean under Two-Phase Sampling Scheme. *Journal of Modern Applied Statistical Methods*, *19*(1), 1-16.

ประวัติผู้เขียน

ชื่อ	อธิปกรณ์ นาถมทอง	
วันเกิด	08 กรกฎาคม 2543	
สถานที่เกิด	โรงพยาบาลกมลาไสย	
สถานที่อยู่ปัจจุบัน	203 หมู่11 ตำบลโคกสะอาด อำเภอฆ้องชัย จังหวัดกาฬสินธุ์ 46130	
ตำแหน่งหน้าที่การงาน	นิสิตระดับบัณฑิตศึกษา	
สถานที่ทำงานปัจจุบัน	มหาวิทยาลัยมหาสารคาม	
ประวัติการศึกษา	พ.ศ. 2564	วิทยาศาสตรบัณฑิต (วท.บ.)
		สาขา สถิติ
		คณะวิทยาศาสตร์
		มหาวิทยาลัยมหาสารคาม
	พ.ศ. 2566	วิทยาศาสตรมหาบัณฑิต (วท.ม.)
		สาขา วิทยาการจัดการสถิติ
		คณะวิทยาศาสตร์
		มหาวิทยาลัยมหาสารคาม