

การเปรียบเทียบประสิทธิภาพการสุ่มตัวอย่างแบบเป็นระบบวงกลมแบบใหม่

A Comparison of the Efficiency for New Circular Systematic Sampling

อิศริน วิทย์วิรานนท์ (Isarin Vithyaviranont)^{*}
อำไพ ทองธีรภาพ (Ampai Thongteeraparp)^{**}
เสาวภา ชัยพิทักษ์ (Saowapa Chaipitak)^{***}

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของค่าเฉลี่ยตัวอย่างที่ได้จากวิธีการสุ่มตัวอย่างแบบเป็นระบบแบบวงกลมแบบใหม่ 3 วิธี ได้แก่ วิธี Modified Balanced Circular Systematic System (MBC) วิธี Modified Centered Circular Systematic System (MCC) และวิธี Modified Systematic Sampling (MSS) จากประชากรที่มีแนวโน้มเชิงเส้น และเปรียบเทียบประสิทธิภาพของวิธีการสุ่มตัวอย่างทั้ง 3 วิธี ด้วยค่าเฉลี่ยคลาดเคลื่อนกำลังสอง (Mean Square Error : MSE) ผลการศึกษาพบว่า เมื่อขนาดประชากรเป็นจำนวนเลขคู่และขนาดตัวอย่างเป็นจำนวนเลขคี่ วิธี MBC และวิธี MCC มีประสิทธิภาพในการประมาณค่าเฉลี่ยใกล้เคียงกัน ส่วนวิธี MSS มีประสิทธิภาพต่ำกว่าวิธี MBC และวิธี MCC

คำสำคัญ : เปรียบเทียบประสิทธิภาพของวิธีการสุ่มตัวอย่าง, Modified Balanced Circular Systematic System, Modified Centered Circular Systematic System, Modified Systematic Sampling

^{*} นิสิตปริญญาโท ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยเกษตรศาสตร์ บางเขน Email: isarin.vith@gmail.com
(Student, Department of Statistics, Faculty of Science, Kasetsart University, Bangkok.)
Email: isarin.vith@gmail.com

^{**} อาจารย์ที่ปรึกษาหลัก, ผู้ช่วยศาสตราจารย์; ประจำภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยเกษตรศาสตร์ บางเขน
(Major Advisor, Assistance Professor; Department of Statistics, Faculty of Science, Kasetsart University, Bangkok.)

^{***} อาจารย์ที่ปรึกษาร่วม, อาจารย์; ประจำภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยเกษตรศาสตร์ บางเขน
(Co-Advisor, Lecturer; Department of Statistics, Faculty of Science, Kasetsart University, Bangkok.)

Abstract

The purposes of this research were to study and compare the efficiency of sample means from three new circular systematic samplings. They composed of the Modified Balanced Circular Systematic System (MBC), Modified Centered Circular Systematic System (MCC) and Modified Systematic Sampling (MSS), The studied population were linear trend. The efficiency of three sampling methods were compared by using the mean square error (MSE). The result showed that when population size was even number and the sample size was odd number, the MBC and the MCC methods provided equally efficacy. The MSS method was less efficiencies than the MBC and the MCC for all situations.

Keywords: Circular Systematic Sampling, Modified Balanced Circular Systematic System, Modified Centered Circular Systematic System, Modified Systematic Sampling

บทนำ

การสุ่มตัวอย่างแบบเป็นระบบเป็นหนึ่งในวิธีการสุ่มตัวอย่างที่ได้รับความนิยมมาก การสุ่มตัวอย่างแบบเป็นระบบแบ่งเป็น 2 แบบ คือการสุ่มตัวอย่างแบบเป็นระบบเชิงเส้นและการสุ่มตัวอย่างแบบเป็นระบบวงกลม การสุ่มตัวอย่างแบบเป็นระบบเชิงเส้น (Linear Systematic Sampling) ใช้ในกรณีที่ ขนาดประชากร (N) มีค่าเท่ากับผลคูณของขนาดตัวอย่าง (n) กับช่วงที่ใช้ในการสุ่ม (k) กล่าวคือ $N = nk$ แต่ในกรณีที่ $N \neq nk$ การสุ่มแบบเป็นระบบเชิงเส้นทำให้ในแต่ละชุดตัวอย่างได้หน่วยตัวอย่างไม่ครบ เช่น กำหนด $N = 20$ $n = 7$ และ $k = 3$ สมมติให้จุดเริ่มต้นเป็น 3 จะทำการสุ่มด้วยช่วงการสุ่มตัวอย่างเท่ากับ 3 จนครบขนาดตัวอย่างที่ต้องการได้ตัวอย่างเป็น 3, 6, 9, 12, 15, 18 และ 21 แต่ตำแหน่งตัวอย่างที่ 21 นั้นไม่มีเนื่องจาก $N = 20$ ทำให้ได้ขนาดตัวอย่างเพียง 6 หน่วย นักสถิติจึงพัฒนาวิธีการสุ่มตัวอย่างแบบเป็นระบบวงกลม (Circular Systematic Sampling) ซึ่งเป็นการสุ่มโดยทำการกระจายตัวอย่างให้อยู่ในลักษณะวงกลมแล้วสุ่มตัวอย่างแบบวนไปเรื่อยๆ เพื่อแก้ไขปัญหาการสุ่มตัวอย่างในกรณีที่ $N \neq nk$

Leu และ Kao (2006) เสนอวิธีการสุ่มตัวอย่างแบบมีระบบวงกลมแบบใหม่ 2 วิธี สามารถใช้ในกรณีที่ $N \neq nk$ ได้แก่ วิธี Modified Balanced Circular Systematic System (MBC) และวิธี Modified Centered Circular Systematic System (MCC) โดยทั้ง 2 วิธีนี้ถูกพัฒนามาจาก Balanced Circular Systematic System และ Centered Circular Systematic System ที่มีข้อจำกัดว่าขนาดของประชากรและขนาดตัวอย่างควรเป็นจำนวนเลขคู่ คิดค้นโดย Uthayakumaran (1998) โดยทั้ง 2 วิธีใหม่นี้ถูกกล่าวว่าเป็นวิธีที่มีประสิทธิภาพมากกว่า Circular Systematic Sampling และเหมาะที่จะใช้เมื่อประชากรมีลักษณะแบบเชิงเส้นหรือแบบเส้นโค้งพาราโบลา Zaheen K. Javid S. และ Sat G. (2013) พัฒนาการสุ่มตัวอย่างแบบเป็นระบบวงกลมแบบใหม่เพื่อเป็นทางเลือกในการสุ่มตัวอย่าง โดยแนวคิดในการพัฒนาคือ แก้ไขปัญหาการเกิดซ้ำของหน่วยตัวอย่างซึ่งเป็นปัญหาหลักของวิธี Circular Systematic Sampling

งานวิจัยนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของค่าเฉลี่ยตัวอย่างที่ได้จากการสุ่มตัวอย่างแบบเป็นระบบวงกลมแบบใหม่ประกอบด้วย วิธี Modified Balanced Circular Systematic System (MBC) วิธี Modified Centered Circular Systematic System (MCC) และวิธี Modified Systematic Sampling (MSS) โดยจะพิจารณาจากค่าเฉลี่ยคลาดเคลื่อนกำลังสอง (Mean Square Error : MSE)

ทฤษฎีที่เกี่ยวข้อง

1. Modified Balanced Circular Systematic System (MBC)

วิธี MBC เป็นวิธีการสุ่มตัวอย่างที่เหมาะสมสำหรับประชากรที่มีแนวโน้มเชิงเส้นหรือแบบแนวโน้มพาราโบริก (Leu & Kao, 2006) โดยมีขั้นตอนการสุ่มตัวอย่างดังนี้

- 1) กำหนดขนาดของประชากร (N) และขนาดของตัวอย่าง (n)
- 2) คำนวณช่วงการสุ่ม $k = \text{INT}\left(\frac{N}{n}\right)$ ถ้าเหลือเศษให้ตัดทิ้ง โดยที่ INT คือ Integer หรือเลข

จำนวนเต็ม

- 3) เลือกจุดสุ่มเริ่มต้น i โดยที่ $1 \leq i \leq N$
- 4) พิจารณา n เป็นจำนวนคู่หรือจำนวนคี่

ในกรณีที่ n เป็นจำนวนคู่ หน่วยตัวอย่าง n หน่วยที่ได้ จะประกอบด้วยหน่วยตัวอย่างที่กำหนดโดย label 1 และ label 2 เมื่อ

$$\begin{aligned} \text{label1} &= i + 2(j-1)k & ; 1 \leq i + 2(j-1)k \leq N \\ &= i + 2(j-1)k - N & ; N < i + 2(j-1)k \quad \text{โดยที่ } j = 1, 2, \dots, \frac{n}{2} \\ \text{label2} &= N - i - 2(j-1)k + 1 & ; 1 \leq i + 2(j-1)k - 1 < N \\ &= 2N - i - 2(j-1)k + 1 & ; N \leq i + 2(j-1)k - 1 \quad \text{โดยที่ } j = 1, 2, \dots, \frac{n}{2} \end{aligned}$$

ในกรณีที่ n เป็นจำนวนคี่ หน่วยตัวอย่าง $n - 1$ หน่วยที่ได้ จะประกอบด้วยหน่วยตัวอย่างที่กำหนดโดย label1 และ label2 เมื่อ

$$\begin{aligned} \text{label1} &= i + 2(j-1)k & ; 1 \leq i + 2(j-1)k \leq N \\ &= i + 2(j-1)k - N & ; N < i + 2(j-1)k \quad \text{โดยที่ } j = 1, 2, \dots, \frac{(n-1)}{2} \\ \text{label2} &= N - i - 2(j-1)k + 1 & ; 1 \leq i + 2(j-1)k - 1 < N \\ &= 2N - i - 2(j-1)k + 1 & ; N \leq i + 2(j-1)k - 1 \quad \text{โดยที่ } j = 1, 2, \dots, \frac{(n-1)}{2} \end{aligned}$$

หน่วยตัวอย่างสุดท้ายกำหนด

$$\text{label} = \frac{(N+1)}{2} \quad \text{ถ้า } N \text{ เป็นจำนวนคี่}$$

$$\text{label} = \frac{N}{2} \quad \text{ถ้า } N \text{ เป็นจำนวนคู่}$$

2. Modified Centered Circular Systematic System (MCC)

วิธี MCC ถูกพัฒนาขึ้นมาพร้อมกับวิธี MBC ซึ่งเป็นวิธีที่เหมาะสมกับกรณีที่มีประชากรมีเชิงเส้นหรือมีแนวโน้มแบบพาราโบริก (Leu & Kao, 2006) โดยมีขั้นตอนการสุ่มตัวอย่าง ดังนี้

- 1) กำหนดขนาดของประชากร (N) และขนาดของตัวอย่าง (n)
- 2) คำนวณช่วงการสุ่ม $k = \text{INT}\left(\frac{N}{n}\right)$ ถ้าเหลือเศษให้ตัดทิ้ง โดยที่ INT คือ Integer หรือเลข

จำนวนเต็ม

- 3) เลือกจุดสุ่มเริ่มต้น i โดยพิจารณาจาก N ว่าเป็นจำนวนคู่หรือจำนวนคี่

ในกรณี N เป็นจำนวนคี่ $i = \frac{(N+1)}{2}$ และในกรณีที่ N เป็นจำนวนคู่ $i = \frac{N}{2}$

4) พิจารณาว่า n เป็นจำนวนคี่หรือจำนวนคู่

ในกรณีที่ n เป็นจำนวนคี่ หน่วยตัวอย่าง $n-1$ หน่วยที่ได้ จะประกอบด้วยหน่วยตัวอย่างที่กำหนด โดย label1 และ label2 เมื่อ

$$\text{label1} = i + (j - 1)k$$

$$\text{label2} = i - (j - 1)k \text{ โดยที่ } j = 1, 2, \dots, \frac{n}{2}$$

$$\text{หน่วยตัวอย่างสุดท้ายกำหนดโดย label } x \text{ โดยที่ } \text{label } x = i + \frac{nk}{2}$$

ในกรณีที่ n เป็นจำนวนคู่ หน่วยตัวอย่าง n หน่วยที่ได้ จะประกอบด้วยหน่วยตัวอย่างที่กำหนด โดย label1 และ label2 เมื่อ

$$\text{label } 1 = i + (j - 1)k$$

$$\text{label } 2 = i - (j - 1)k \text{ โดยที่ } j = 1, 2, \dots, \frac{(n+1)}{2}$$

สำหรับการคำนวณค่าเฉลี่ยตัวอย่างเมื่อสุ่มตัวอย่างด้วยวิธี MBC นั้นต้องพิจารณาในกรณีที่ N เป็นจำนวนคี่ หรือ n เป็นจำนวนคู่ และสำหรับการสุ่มตัวอย่างวิธี MCC ที่มี N และ n เป็นจำนวนคี่ สามารถคำนวณค่าเฉลี่ยตัวอย่าง โดย

$$\bar{y} = \frac{\sum_{i=1}^n y_i}{n}$$

โดยที่ y_i คือ ค่าสังเกตที่ได้จากหน่วยตัวอย่างที่ i ; $i = 1, 2, \dots, n$

n คือ ขนาดตัวอย่าง

สำหรับกรณี N และ n ไม่เป็นไปตามที่กล่าวข้างต้น วิธี MBC และ MCC จะคำนวณค่าเฉลี่ยตัวอย่างด้วยสูตร Corrected Sample Mean ดังนี้

$$\bar{y}_{cl} = \frac{1}{n} \left\{ S_{11}y_{x_1} + \sum_{j=2}^{n-1} y_{x_j} + S_{12}y_{x_n} \right\}$$

โดยที่ y_{x_j} คือ ค่าของหน่วยตัวอย่างลำดับที่ x_j ; $j = 1, 2, \dots, n$

x_j คือ ลำดับของหน่วยตัวอย่างที่สุ่มมาเป็นตัวอย่าง

$$S_{11} = \frac{E_1 - 2x_n}{x_1 - x_n}$$

$$S_{12} = \frac{2x_1 - E_1}{x_1 - x_n}$$

$$E_1 = \frac{n(N+1)}{2} - \sum_{j=2}^{n-1} x_j$$

3. Modified Systematic Sampling (MSS)

เป็นวิธีการสุ่มแบบเป็นระบบวงกลม สามารถใช้ได้ทั้งกรณี $N = nk$ และ $N \neq nk$ และแก้ปัญหาการเกิดหน่วยตัวอย่างซ้ำของการสุ่มแบบ Circular Systematic Sampling อีกด้วย (Zaheen, 2013) โดยในการสุ่มตัวอย่างแบบ MSS มีขั้นตอนการสุ่มดังนี้

- 1) หาตัวคูณร่วมน้อยของ N กับ n เรียกว่าค่า L
- 2) หาค่า k^* , m , s , k ได้ โดยที่ $k^* = \frac{L}{n}$, $m = \frac{L}{N}$, $s = \frac{N}{k^*}$, $k = \frac{k^*}{m}$ และ $ms = n$
- 3) เลือกตำแหน่งการสุ่มโดย $y_{k^* + (m-1)k + (s-1)k^*}$

คำนวณค่าเฉลี่ยตัวอย่างของวิธี MSS โดย

$$\bar{y}_{MSS} = \frac{\sum_{j=0}^{m-1} \sum_{t=0}^{s-1} y_{r+jk+tk^*}}{ms}$$

ตัวอย่างการสุ่มด้วยวิธี MBC

กำหนดให้ประชากรมีแนวโน้มเชิงเส้น $N = 20$ หน่วยประกอบด้วย 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41 กำหนด $n = 7$ เลือกตัวอย่างด้วยวิธี MBC

1. ให้ลำดับแก่ข้อมูลหน่วยประชากร 1 ถึง 20
2. คำนวณค่า $k = \frac{20}{7} = 2.857 = 2$ และเลือกจุดเริ่มต้นการสุ่ม (i) สมมติให้ $i = 3$
3. พิจารณา n ว่าเป็นจำนวนคู่หรือจำนวนคี่ จากนั้นพิจารณา label1 และ label2 ได้หน่วยตัวอย่างดังนี้

$$\text{label1} = i + 2(j-1)k \quad ; 1 \leq i + 2(j-1)k \leq N$$

$$i + 2(j-1)k - N \quad ; N < i + 2(j-1)k \quad \text{โดยที่ } j = 1, 2, \dots, \frac{(n-1)}{2}$$

$$\text{label2} = N - i - 2(j-1)k + 1 \quad ; 1 \leq i + 2(j-1)k - 1 < N$$

$$2N - i - 2(j-1)k + 1 \quad ; N \leq i + 2(j-1)k - 1 \quad \text{โดยที่ } j = 1, 2, \dots, \frac{(n-1)}{2}$$

จาก label1 พิจารณา $i + 2(j-1)k$ ที่ค่า $j = 1, 2, 3$ ได้ผลดังนี้

$$4 + 2(1-1)2 = 4 \quad ; 1 \leq 4 \leq 20$$

$$4 + 2(2-1)2 = 8 \quad ; 1 \leq 8 \leq 20$$

$$4 + 2(3-1)2 = 12 \quad ; 1 \leq 12 \leq 20$$

ดังนั้น จากการพิจารณา label1 เมื่อ $j = 1, 2, 3$ พบว่าทั้ง 3 ค่าที่ได้ตกอยู่ในช่วง

$1 \leq i + 2(j-1)k \leq N$ ทำให้สรุปได้ว่า label1 ให้ตำแหน่งหน่วยตัวอย่างเป็น 4, 8, 12

จาก label2 พิจารณา $i + 2(j-1)k - 1$ ที่ค่า $j = 1, 2, 3$ ได้ผลดังนี้

$$4 + 2(1-1)2 - 1 = 3 \quad ; 1 \leq 3 < 20$$

$$4 + 2(2-1)2 - 1 = 7 \quad ; 1 \leq 7 < 20$$

$$4 + 2(3-1)2 - 1 = 11 \quad ; 1 \leq 11 < 20$$

ดังนั้น จากการพิจารณา label2 เมื่อ $j = 1, 2, 3$ พบว่าทั้ง 3 ค่าที่ได้ตกอยู่ในช่วง $1 \leq i + 2(j - 1)k \leq N$ ดังนั้น label2 จะได้หน่วยตัวอย่างดังนี้

$$20 - 4 - 2(1 - 1)2 + 1 = 17$$

$$20 - 4 - 2(2 - 1)2 + 1 = 13$$

$$20 - 4 - 2(3 - 1)2 + 1 = 9$$

ทำให้สรุปได้ว่า label2 ให้ตำแหน่งหน่วยตัวอย่างเป็น 9, 13, 17 และ label สุดท้ายใช้ตำแหน่งกลางของประชากรเป็นหน่วยตัวอย่างสุดท้ายคือ $\text{label} = \frac{20}{2} = 10$ ดังนั้นสรุปตำแหน่งหน่วยตัวอย่างที่ได้ทั้งหมดจากวิธีการสุ่ม MBC ดังนี้ 4, 8, 9, 10, 12, 13, 17 หรือได้หน่วยตัวอย่างเป็น 9, 17, 19, 21, 25, 27, 35

คำนวณค่าเฉลี่ยตัวอย่างของวิธี MBC ในกรณี N เป็นจำนวนคู่ และ n เป็นจำนวนคี่ อยู่ในกรณีที่ใช้ $\bar{y}_{cl}$ หรือ Corrected Sample mean มีขั้นตอนดังนี้

1. พิจารณารูปแบบเชิงเส้นของประชากร อ้างอิงจากข้อมูลประชากรที่กำหนดไว้ พบว่ามีประชากรมีรูปแบบเชิงเส้นคือ $y_i = 1 + 2i$ โดยที่ i คือลำดับของหน่วยประชากร

2. เนื่องจากประชากรมีรูปแบบ $y_i = 1 + 2i$ และหน่วยตัวอย่างที่ได้จากการสุ่ม 7 หน่วย ได้แก่ $x_1 = 9, x_2 = 17, x_3 = 19, x_4 = 21, x_5 = 25, x_6 = 27$ และ $x_7 = 35$ ดังนั้นคำนวณค่า y_{x_j} ได้ค่าต่างๆดังนี้

$$y_{x_1} = 1 + 2(9) = 19$$

$$y_{x_2} = 1 + 2(17) = 35$$

$$y_{x_3} = 1 + 2(19) = 39$$

$$y_{x_4} = 1 + 2(21) = 43$$

$$y_{x_5} = 1 + 2(25) = 51$$

$$y_{x_6} = 1 + 2(27) = 55$$

$$y_{x_7} = 1 + 2(35) = 71$$

$$E_1 = \frac{n(N+1)}{2} - \sum_{j=2}^6 x_j = \frac{7(20+1)}{2} - 109 = -35.5$$

$$S_{11} = \frac{E_1 - 2x_n}{x_1 - x_n} = \frac{-35.5 - 2(35)}{9 - 35} = 4.06$$

$$S_{12} = \frac{2x_1 - E_1}{x_1 - x_n} = \frac{2(9) + 35.5}{9 - 35} = -2.06$$

$$\bar{y}_{cl} = \frac{1}{n} \left\{ S_{11}y_{x_1} + \sum_{j=2}^{n-1} y_{x_j} + S_{12}y_{x_n} \right\} = \frac{1}{7} \{ 4.06(19) + 223 - 2.06(71) \} = 22$$

ดังนั้นค่าเฉลี่ยที่คำนวณจาก $\bar{y}_{cl} = 22$ ซึ่งได้ค่าเฉลี่ยตัวอย่างเท่ากับค่าเฉลี่ยประชากรโดยที่

$$\bar{Y} = \frac{3 + 5 + \dots + 41}{20} = 22$$

ตัวอย่างการสุ่มด้วยวิธี MCC

กำหนดให้ประชากรมีแนวโน้มเชิงเส้น $N = 20$ ประกอบด้วย 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41 กำหนด $n = 7$ เลือกตัวอย่างด้วยวิธี MCC

1. ให้ลำดับแก้มข้อมูลหน่วยประชากร 1 ถึง 20

2. คำนวณค่า $k = \frac{20}{7} = 2.857 = 2$ กำหนดจุดเริ่มต้นการสุ่ม (i) โดยใช้ตำแหน่งกลางของ

หน่วยประชากร ดังนั้น $i = \frac{20}{2} = 10$

3. พิจารณา n ว่าเป็นจำนวนคู่หรือจำนวนคี่ จากนั้นพิจารณา label1 และ label2 ได้หน่วยตัวอย่างดังนี้

$$\text{label1} = i + (j - 1)k, \text{ โดยที่ } j = 1, 2, \dots, \frac{(n+1)}{2}$$

$$\text{label2} = i - (j - 1)k, \text{ โดยที่ } j = 1, 2, \dots, \frac{(n+1)}{2}$$

พิจารณา label1 เมื่อ $j = 1, 2, 3, 4$ ได้หน่วยตัวอย่างเป็น

$$10 + (1-1)2 = 10$$

$$10 + (2-1)2 = 12$$

$$10 + (3-1)2 = 14$$

$$10 + (4-1)2 = 16$$

ดังนั้น จากการพิจารณา label1 เมื่อ $j = 1, 2, 3, 4$ พบว่าตำแหน่งหน่วยตัวอย่างได้แก่ 10, 12, 14, 16

พิจารณา label2 เมื่อ $j = 1, 2, 3, 4$ ได้หน่วยตัวอย่างเป็น

$$10 - (1-1)2 = 10$$

$$10 - (2-1)2 = 8$$

$$10 - (3-1)2 = 6$$

$$10 - (4-1)2 = 4$$

ดังนั้น จากการพิจารณา label1 เมื่อ $j = 1, 2, 3, 4$ พบว่าตำแหน่งหน่วยตัวอย่างได้แก่ 4, 8, 6, 10 แต่เนื่องจาก เมื่อ $j = 1$ พบว่าตำแหน่งที่ได้คือจุดเริ่มต้นของการสุ่มจึงไม่นับเป็นหน่วยตัวอย่าง

สรุปคือตำแหน่งหน่วยตัวอย่างที่ได้จาก label1 และ label2 คือ 4, 6, 8, 10, 12, 14, 16 หรือ หน่วยตัวอย่างคือ 9, 13, 17, 21, 25, 29, 33

ตัวอย่างการสุ่มด้วยวิธี MSS

กำหนดให้ประชากรมีแนวโน้มเชิงเส้นขนาดเป็น $N = 20$ ประกอบไปด้วย 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41 ต้องการ $n = 7$ เลือกตัวอย่างด้วยวิธี MSS

1. ให้ลำดับแก้มข้อมูลหน่วยประชากร 1 ถึง 20

2. หาผลคูณร่วมน้อยของ N และ n หรือ $L = (20, 7) = 140$ จากนั้นคำนวณค่าต่างๆ ดังนี้

$$k^* = \frac{L}{n} = \frac{140}{7} = 20$$

$$m = \frac{L}{N} = \frac{140}{20} = 7$$

$$s = \frac{N}{k^*} = \frac{20}{20} = 1$$

$$k = \text{Round}\left(\frac{20}{7}\right) = 3$$

4. นำค่าที่ได้มาคำนวณหาตำแหน่งตัวอย่างที่จะสุ่มด้วยสูตร $y_{k^* + (m-1)k + (s-1)k^*}$ ได้ชุดหน่วยตัวอย่างทั้งหมดที่สามารถเกิดขึ้นได้ดังนี้

m	s	k* = 1	k* = 2	k* = 3	...	k* = 10	...	k* = 20
1	1	1	2	3	...	10	...	20
2	1	4	5	6	...	13	...	3
3	1	7	8	9	...	16	...	6
4	1	10	11	12	...	19	...	9
5	1	13	14	15	...	2	...	12
6	1	16	17	18	...	5	...	15
7	1	19	20	1	...	8	...	18

กล่าวได้ว่า k^* เป็นค่าที่จำนวนชุดข้อมูล ในกรณีนี้ $k^* = 20$ แสดงว่ามีจำนวนชุดหน่วยตัวอย่างจากการสุ่มด้วยวิธี MSS ทั้งหมด 20 ชุด ดังนั้นเมื่อแทนค่าต่างๆ ลงในสมการ $y_{k^* + (m-1)k + (s-1)k^*}$ แล้วจึงได้หน่วยตัวอย่างที่เป็นไปได้จำนวน 20 ชุด จากนั้นเลือกมาเพียงชุดเดียว

เช่น เมื่อหากสุ่มได้ชุด $k^* = 10$ จะได้ตำแหน่งหน่วยตัวอย่างเป็น 2, 5, 8, 10, 13, 16, 19 หรือหน่วยตัวอย่างคือ 5, 11, 17, 21, 27, 33, 39

วิธีการวิจัย

1. จำลองข้อมูลประชากรด้วยตัวแบบอภิประชากร (Super-population model) ที่มีรูปแบบเชิงเส้น คือ $y_i = \alpha + \beta + e_i$ กำหนดให้ α และ β เท่ากับ 7 และ 10 ตามลำดับ กำหนดให้ $E(e_i) = 0$, $\text{Var}(e_i) = i^2$ เมื่อ g คือค่าคงที่กำหนด (Predetermined constant) โดยกำหนด $g = 0, 1$ และ 2 เนื่องจากผู้วิจัยต้องการทดสอบว่าในกรณีที่ข้อมูลมีความแปรปรวนมากขึ้น ประสิทธิภาพในการประมาณค่าเฉลี่ยจะเปลี่ยนไปอย่างไร

2. กำหนดขนาดประชากร (N) และขนาดตัวอย่าง (n) โดยใช้เกณฑ์ดังนี้ (1) เมื่อ $N \geq 1,000$ กำหนด n เท่ากับ 10 และ 15% ของขนาดประชากร (2) เมื่อ $N < 1,000$ กำหนด n เท่ากับ 15 และ 30% ของขนาดประชากร

ตารางที่ 1 สรุปรูปขนาดประชากรและขนาดตัวอย่าง

N	n	%
1500	225	15%
	150	10%
1100	165	15%
	110	10%
700	210	30%
	105	15%
500	150	30%
	75	15%
300	90	30%
	45	15%

3. สุ่มตัวอย่างและคำนวณค่าเฉลี่ยตัวอย่างจากวิธีการสุ่มตัวอย่าง 3 วิธี

4 นำชุดตัวอย่างที่ได้จากวิธีการสุ่มตัวอย่างทั้ง 3 วิธีมาคำนวณค่า MSE ดังนี้

$$MSE(\bar{y}) = \frac{\sum_{i=1}^R (\bar{y}_i - \mu)^2}{R}$$

โดยที่ $\bar{y}_i$ คือ ค่าเฉลี่ยตัวอย่างจากวิธีการสุ่มตัวอย่างแต่ละครั้ง โดยที่ $i = 1, 2, \dots, 1,000$

μ คือ ค่าเฉลี่ยประชากร

R คือ จำนวนครั้งในการทำการทดลอง โดยกระทำซ้ำ 1,000 ครั้ง

นำค่า MSE ที่ได้มาคำนวณค่าประสิทธิภาพเชิงสัมพัทธ์ (Relative Efficiency) เพื่อเปรียบเทียบประสิทธิภาพของวิธีการสุ่มตัวอย่างแบบเป็นระบบวงกลมทั้ง 3 วิธี ดังนี้

$$RE = \frac{MSE(\bar{y}_A)}{MSE(\bar{y}_B)}$$

ถ้า $RE = 1$ แสดงว่า $MSE(\bar{y}_A)$ มีค่าเท่ากับ $MSE(\bar{y}_B)$ สรุปได้ว่า $\bar{y}_A$ และ $\bar{y}_B$ มีประสิทธิภาพเท่ากัน

ถ้า $RE < 1$ แสดงว่า $MSE(\bar{y}_A)$ มีค่าต่ำกว่า $MSE(\bar{y}_B)$ สรุปได้ว่า $\bar{y}_A$ มีประสิทธิภาพสูงกว่า $\bar{y}_B$

ถ้า $RE > 1$ แสดงว่า $MSE(\bar{y}_A)$ มีค่าสูงกว่า $MSE(\bar{y}_B)$ สรุปได้ว่า $\bar{y}_A$ มีประสิทธิภาพต่ำกว่า $\bar{y}_B$

ผลการวิจัย

จากการเปรียบเทียบค่าเฉลี่ยตัวอย่างที่คำนวณได้จากวิธีการสุ่มตัวอย่างแบบเป็นระบบแบบ MBC MCC และ MSS ค่า MSE และค่า RE แสดงดังตารางที่ 2 และตารางที่ 3

ตารางที่ 2 MSE ของค่าเฉลี่ยตัวอย่างของ MBC MCC และ MSS

N	n	g	$MSE(\bar{y}_{MBC})$	$MSE(\bar{y}_{MCC})$	$MSE(\bar{y}_{MSS})$
1500	225	g = 0	0.000687	0.000662	359.92136
		g = 1	0.503948	0.463824	370.55103
		g = 2	515.55282	485.3774	4336.5995
	150	g = 0	0.007599	0.000662	845.4335
		g = 1	5.663915	0.530809	808.01887
		g = 2	5711.5294	529.2961	6373.3817
1100	165	g = 0	0.00093	0.000928	374.2759
		g = 1	0.530251	0.473273	367.61247
		g = 2	359.22675	383.50325	3168.5909
	110	g = 0	0.010189	0.000861	798.09099
		g = 1	5.406561	0.491321	799.53038
		g = 2	3717.3429	358.72178	4747.3334
N	n	g	$MSE(\bar{y}_{MBC})$	$MSE(\bar{y}_{MCC})$	$MSE(\bar{y}_{MSS})$
700	210	g = 0	0.006321	0.001408	93.48419
		g = 1	2.24064	0.001395	92.58033
		g = 2	1030.0475	226.49608	1048.1167
	105	g = 0	0.001404	0.001453	366.49846
		g = 1	0.484616	0.514635	376.81464
		g = 2	227.865	246.25153	2245.0335

ตารางที่ 2 (ต่อ)

N	n	g	MSE($\bar{y}_{MBC}$)	MSE($\bar{y}_{MCC}$)	MSE($\bar{y}_{MSS}$)
500	150	g = 0	0.009084	0.002028	93.31397
		g = 1	2.030978	0.486906	96.52143
		g = 2	754.56586	158.31072	800.48755
	75	g = 0	0.002038	0.001855	358.10576
		g = 1	0.530248	0.515755	395.63636
		g = 2	175.25595	171.42395	1606.6341
300	90	g = 0	0.013903	0.003363	90.06299
		g = 1	2.185539	0.487534	90.15032
		g = 2	443.29759	96.515002	533.38767
	45	g = 0	0.003356	0.003628	357.62785
		g = 1	0.504503	0.5249	364.52894
		g = 2	100.16379	100.17449	1095.9813

ตารางที่ 3 ค่าประสิทธิภาพสัมพัทธ์ (RE) ของวิธี MBC และวิธี MCC เทียบกับวิธี MSS

N	n	g	$\frac{MSE(\bar{y}_{MBC})}{MSE(\bar{y}_{MSS})}$	$\frac{MSE(\bar{y}_{MCC})}{MSE(\bar{y}_{MSS})}$
1500	225	g = 0	1.91E-06	1.84E-06
		g = 1	0.00136	0.0012517
		g = 2	0.1188841	0.1119258
	150	g = 0	8.99E-06	7.84E-07
		g = 1	0.0070096	0.0006569
		g = 2	0.8961537	0.0830479

ตารางที่ 3 (ต่อ)

N	n	g	$\frac{MSE(\bar{y}_{MBC})}{MSE(\bar{y}_{MSS})}$	$\frac{MSE(\bar{y}_{MCC})}{MSE(\bar{y}_{MSS})}$
1100	165	g = 0	2.48E-06	2.48E-06
		g = 1	0.0014424	0.0012874
		g = 2	0.1133711	0.1210327
	110	g = 0	1.28E-05	1.08E-06
		g = 1	0.0067622	0.0006145
		g = 2	0.7830381	0.0755628
700	210	g = 0	6.76E-05	1.51E-05
		g = 1	0.0242021	1.51E-05
		g = 2	0.9827603	0.2160982
	105	g = 0	3.83E-06	3.97E-06
		g = 1	0.0012861	0.0013658
		g = 2	0.1014974	0.1096872
500	150	g = 0	9.73E-05	2.17E-05
		g = 1	0.0210417	0.0050445
		g = 2	0.9426328	0.1977679
	75	g = 0	5.69E-06	5.18E-06
		g = 1	0.0013402	0.0013036
		g = 2	0.1090827	0.1066976
300	90	g = 0	0.0001544	3.73E-05
		g = 1	0.0242433	0.005408
		g = 2	0.8310983	0.1809472
	45	g = 0	9.38E-06	1.01E-05
		g = 1	0.001384	0.0014399
		g = 2	0.0913919	0.0914016

อภิปรายผล

1. ค่า MSE ของ MBC และ MCC ให้ค่าใกล้ 0 มากในกรณีที่ขนาดประชากรเป็นจำนวนเลขคู่และขนาดตัวอย่างเป็นจำนวนเลขคี่ กล่าวคือ การประมาณค่าเฉลี่ยตัวอย่างด้วย Corrected Sample Mean ให้ค่าใกล้เคียงค่าเฉลี่ยประชากรมาก
2. เมื่อ $g = 0$ และ 1 ค่า MSE ของวิธี MBC และ MCC มีค่าเข้าใกล้ 0 สำหรับกรณีขนาดประชากรเป็นจำนวนคู่และขนาดตัวอย่างเป็นจำนวนคี่ กล่าวคือมีประสิทธิภาพการประมาณค่าเฉลี่ยสูง กรณีขนาดประชากรและขนาดตัวอย่างอื่นๆ วิธี MBC ให้ค่า MSE มากขึ้นแต่ไม่ห่างจาก 0 มาก สำหรับวิธี MSS ให้ค่า MSE ออกจาก 0 อย่างมากเมื่อเทียบกับวิธี MBC และ MCC
3. กรณีที่ $g = 2$ ค่า MSE ของทุกวิธีการสุ่มตัวอย่างมีค่าออกจาก 0 มากเมื่อเทียบกับกรณีที่ g เท่ากับ 0 และ 1 กล่าวคือเมื่อ g มีค่ามากขึ้นข้อมูลประชากรจะมีความแปรปรวนมากขึ้นส่งผลให้การประมาณค่าเฉลี่ยตัวอย่างจะมีความแม่นยำน้อยลงอย่างมาก
4. จากผลการศึกษาการเปรียบเทียบประสิทธิภาพการประมาณค่าเฉลี่ยของการสุ่มตัวอย่างแบบเป็นระบบวงกลมใหม่ทั้ง 3 วิธี ได้ข้อสรุปว่าวิธี MCC มีประสิทธิภาพดีที่สุดในทุกกรณี แต่สำหรับกรณีที่ขนาดประชากรและขนาดตัวอย่างเป็นจำนวนเลขคู่และเลขคี่ตามลำดับแล้ว ค่า MSE ที่ได้จะมีค่าใกล้เคียงกับวิธี MBC สำหรับวิธี MSS ให้ MSE มากกว่าทั้ง 2 วิธี กล่าวได้ว่าวิธี MSS มีประสิทธิภาพในการประมาณค่าเฉลี่ยต่ำกว่าวิธี MBC และวิธี MCC
5. จากการศึกษาพบว่าวิธีการคำนวณค่าเฉลี่ยตัวอย่างด้วย Corrected Sample Mean จะให้ค่าเฉลี่ยตัวอย่างที่เข้าใกล้หรือเท่ากับค่าเฉลี่ยประชากร สามารถสังเกตจากค่า MSE ที่เข้าใกล้ 0 ในกรณีที่ g เท่ากับ 0 และ 1 กล่าวคือสามารถประมาณค่าเฉลี่ยตัวอย่างได้แม่นยำ แต่สำหรับกรณีที่ $g = 2$ การประมาณค่าเฉลี่ยด้วยวิธีนี้ให้ค่าผิดเพี้ยนไปมากเช่นกัน แต่ยังไม่มากเท่าวิธีคำนวณค่าเฉลี่ยตัวอย่างทั่วไป

ข้อเสนอแนะ

1. ในกรณีที่ข้อมูลมีรูปแบบเชิงเส้นแนะนำวิธี MBC และวิธี MCC มีประสิทธิภาพใกล้เคียงกันและทั้ง 2 วิธีมีประสิทธิภาพในการประมาณค่าเฉลี่ยดีกว่าวิธี MSS
2. วิธี MSS จะทำให้หน่วยตัวอย่างที่ได้เกิดซ้ำได้ยากมาก

กิตติกรรมประกาศ

งานวิจัยชิ้นนี้สำเร็จลงไปได้ด้วยดีด้วยความช่วยเหลือจากผู้ช่วยศาสตราจารย์ ดร. อำไพ ทองธิภาพ อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก และดร. เสาวภา ชัยพิทักษ์ อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม ที่ได้กรุณาให้คำปรึกษาและคำแนะนำในการทำงานวิจัยชิ้นนี้ ตลอดจนตรวจแก้ไขข้อบกพร่องต่างๆ จนกระทั่งงานวิจัยชิ้นนี้เสร็จสมบูรณ์ ผู้วิจัยขอกราบขอบพระคุณและสำนึกในความกรุณาเป็นอย่างสูง

เอกสารอ้างอิง

ภาษาไทย

- นครินทร์ แก้วสระแสน. 2555. “การเปรียบเทียบประสิทธิภาพการสุ่มตัวอย่างแบบเป็นระบบ 4 วิธี.” **Veridian E-Journal, SU** 2555, ฉบับที่ 5 พฤศจิกายน – ธันวาคม : 420 – 428.
- สุชาดา กิระนันท์. 2538. **ทฤษฎีและวิธีการสำรวจตัวอย่าง**. พิมพ์ครั้งที่ 1. โรงพิมพ์จุฬาลงกรณ์มหาวิทยาลัย, กรุงเทพฯ.

ภาษาต่างประเทศ

- Leu, C.K. and F.F. Kao. 2006. Modified Balanced Circular Systematic Sampling. **Statistics & Probability Letters**. 76: 373-383.
- Sampath S. and Valarakshmi V.. 2008. Diagonal Circular Systematic Sampling. **Model Assisted Statistics and Applications** 3. 345-352.
- Uthayakumaran N., 1998. Additional Circular Systematic Sampling Method. **Biometrical Journal**. 40: 467-474.
- Zaheen K., Javid S. and Sat G.. 2013. A New Sampling Design for Systematic Sampling. **Statistics Theory and Methods**. 42: 3359-3370.