

การเปรียบเทียบประสิทธิภาพการเลือกตัวอย่างแบบมีระบบ 4 วิธี

COMPARATIVE EFFICIENCY OF FOUR SYSTEMATIC SAMPLING METHODS

นรินทร์ แก้วสระแสน*

อำไพ ทองธีรภาพ**

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์ เพื่อเปรียบเทียบประสิทธิภาพของค่าเฉลี่ยตัวอย่างที่ได้จากการเลือกตัวอย่างแบบมีระบบ 4 วิธี คือ การเลือกตัวอย่างแบบมีระบบเชิงเส้น (LSY) การเลือกตัวอย่างแบบมีระบบเชิงเส้นคู่ (BSY) การเลือกตัวอย่างแบบมีระบบชนิด Modified Balanced Circular (MBC) และ การเลือกตัวอย่างแบบมีระบบชนิด Mixed Random (MRSS) เมื่อประชากรมีแนวโน้มเชิงเส้น (Linear Trend) ผลการศึกษาพบว่าค่าเฉลี่ยตัวอย่างที่ได้จากการเลือกตัวอย่างด้วยวิธี MBC ให้ค่า MSE ต่ำที่สุด เมื่อขนาดประชากรเป็นจำนวนคี่ หรือ ขนาดตัวอย่างเป็นจำนวนคู่ แต่เมื่อขนาดประชากรเป็นจำนวนคู่ และขนาดตัวอย่างเป็นจำนวนคี่ การเลือกตัวอย่างด้วยวิธี LSY และ BSY จะให้ค่า MSE ต่ำที่สุด

คำสำคัญ : การเลือกตัวอย่างแบบมีระบบเชิงเส้น แบบมีระบบเชิงเส้นคู่ แบบมีระบบชนิด Modified Balanced Circular แบบมีระบบชนิด Mixed Random

ABSTRACT

The objective of this research is to compare the efficiency of sample means from linear systematic sampling (LSY) , bilinear systematic sampling (BSY) , modified balanced circular systematic sampling (MBC) and mixed random systematic sampling (MRSS) from the population with linear trend. The results show that the sample means from MBC provide the minimum MSE when the population size is odd or the sample size is even. while the population size is even and the sample size is odd the MSE from LSY and BSY is minimum.

KEYWORDS : Linear systematic sampling , Bilinear systematic sampling , Modified balanced circular systematic sampling , Mixed random systematic sampling

บทนำ

การเก็บข้อมูลจากทุกหน่วยของประชากร อาจทำให้เสียเวลานานและมีค่าใช้จ่ายสูงมาก นอกจากนี้ข้อมูลบางประเภทอาจมีข้อจำกัดที่ไม่สามารถเก็บข้อมูลได้จากทุกหน่วยของประชากร การ

* ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยเกษตรศาสตร์ ลาดยาว จตุจักร กทม. 10900 E-mail :first_fasuto@hotmail.com

** ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยเกษตรศาสตร์ ลาดยาว จตุจักร กทม. 10900

เลือกข้อมูลมาเพียงบางส่วนของประชากร เรียกว่า การเลือกตัวอย่าง (Sampling) จะช่วยลดระยะเวลาและค่าใช้จ่ายในการเก็บข้อมูล การเลือกตัวอย่างจึงเป็นส่วนสำคัญของงานวิจัย เทคนิคการเลือกตัวอย่างที่มีประสิทธิภาพจะทำให้ได้ตัวแทนที่ดีจากประชากร วิธีการเลือกตัวอย่างนั้นมีหลายวิธีด้วยกัน ซึ่งการเลือกตัวอย่างแบบมีระบบเชิงเส้น (Linear Systematic Sampling : LSY) เป็นเทคนิคการเลือกตัวอย่างหนึ่งที่ยอดนิยมอย่างแพร่หลาย แต่มีข้อเสียคือ เมื่อขนาดของประชากรหารด้วยขนาดตัวอย่างไม่ลงตัวหรือ $N \neq nk$ โดยที่ ขนาดตัวอย่าง (n) และ ช่วงการสุ่ม (k) เป็นจำนวนเต็ม ทำให้ขนาดตัวอย่างไม่เท่ากับ n หรือไม่เท่ากันและความแปรปรวนของค่าเฉลี่ยเป็นตัวประมาณที่เอนเอียง (Biased Estimator) ซึ่งมีการพัฒนาวิธีการเลือกตัวอย่างแบบมีระบบเพื่อแก้ไขปัญหาดังกล่าวและมีประสิทธิภาพเหมาะสมกับลักษณะของข้อมูล เช่นการเลือกตัวอย่างแบบมีระบบชนิด Mixed Random (MRSS) ซึ่งพัฒนาขึ้นจากการรวมกันระหว่าง การเลือกตัวอย่างแบบง่าย (Sample Random Sampling : SRS) และ LSY แต่มีข้อจำกัดคือการหาค่าความแปรปรวนของตัวอย่าง มีความซับซ้อนและยุ่งยากมากเมื่อเทียบกับการเลือกตัวอย่างแบบมีระบบชนิดอื่นๆ อย่างไรก็ตามยังไม่มียานวิจัยใดทำการเปรียบเทียบประสิทธิภาพของการเลือกตัวอย่างแบบมีระบบชนิด MRSS กับวิธีการเลือกตัวอย่างแบบมีระบบอื่นๆ ผู้วิจัยจึงขอเสนอการเปรียบเทียบประสิทธิภาพระหว่างการเลือกตัวอย่างแบบมีระบบทั้ง 4 ชนิด คือ การเลือกตัวอย่างแบบมีระบบเชิงเส้น (LSY) การเลือกตัวอย่างแบบมีระบบเชิงเส้นคู่ (Bilinear Systematic Sampling : BSY) การเลือกตัวอย่างแบบมีระบบชนิด Modified Balanced Circular (MBC) และ การเลือกตัวอย่างแบบมีระบบชนิด Mixed Random (MRSS) และหาแนวทางการเลือกใช้วิธีการเลือกตัวอย่างแบบมีระบบที่เหมาะสมไปใช้กับข้อมูล

กรอบแนวคิดและทฤษฎี

Huang (2004) นำเสนอ การเลือกตัวอย่างแบบ Mixed Random Systematic Sampling (MRSS) ซึ่งพัฒนามาจากการเลือกตัวอย่างแบบมีระบบชนิดเชิงเส้นและการเลือกตัวอย่างแบบง่าย โดยมีวิธีดำเนินการดังต่อไปนี้

1. เลือกจุดสุ่มเริ่มต้น t โดยที่ $1 \leq t \leq N$
2. แบ่งประชากรออกเป็น 2 กลุ่ม โดยกลุ่มที่ 1 สมาชิกจะเริ่มจาก t ถึง $t + ((n-r)k) - 1$ มีสมาชิกทั้งหมด $(n-r)k$ หน่วย และหน่วยที่เหลือจากการแบ่งกลุ่มที่ 1 ให้เป็นสมาชิกของกลุ่มที่ 2
3. เลือกหน่วยตัวอย่างจากกลุ่มที่ 1 ด้วยการเลือกตัวอย่างแบบง่าย (SRS) มาจำนวน $(n-r)$ หน่วย
4. หน่วย โดยจุดเริ่มต้นสุ่มคือ $t + ((n-r)k) - 1$ จะได้ขนาดตัวอย่าง เท่ากับ n หน่วย

ค่าเฉลี่ยตัวอย่าง คือ

$$\bar{y}_{MRSS} = \frac{1}{N} \sum_{i=1}^n \frac{Y_i}{\left(\frac{n}{N}\right)}$$

$$\text{ความแปรปรวนของตัวอย่าง คือ } \hat{V}(\bar{y}_{MRSS}) = \frac{1}{N^2} \sum_{u_i, u_j \in S_i} \sum_{j=1}^n \left(\frac{1}{\pi_{ij}} - \frac{N^2}{n^2} \right) (y_i - y_j)^2$$

y_i = หน่วยตัวอย่างที่ i เมื่อ $i = 1, 2, \dots, n$

N = ขนาดประชากร

n = ขนาดตัวอย่าง

t = จุดเริ่มสุ่ม

k = ช่วงการสุ่ม

r = ส่วนที่เหลือจาก $\frac{N}{n}$

π_{ij} = ความน่าจะเป็นร่วม (Joint inclusion probabilities)

การเลือกตัวอย่างแบบมีระบบวิธีต่างๆที่นำมาเปรียบเทียบกับในงานวิจัยนี้คำนวณค่าเฉลี่ยของตัวอย่าง ดังนี้

1. การเลือกตัวอย่างแบบมีระบบเชิงเส้น (LSY) (จามรี, 2543)

$$\text{ค่าเฉลี่ยตัวอย่าง คือ } \bar{y}_{LSY} = \frac{1}{n} \sum_{j=1}^n \sum_{i=1}^k y_{ij}$$

y_{ij} = หน่วยตัวอย่างที่ j ในชุดตัวอย่างที่ i เมื่อ $j = 1, 2, \dots, n$ $i = 1, 2, \dots, k$

2. การเลือกตัวอย่างแบบมีระบบเชิงเส้นคู่ (BSY) (ชรินทร์ และ อำนวย, 2540)

$$\text{ค่าเฉลี่ยตัวอย่าง คือ } \bar{y}_{BSY} = \frac{(n-r)k\bar{y}_1 + r(k+1)\bar{y}_2}{nk+r}$$

y_{ij} = หน่วยตัวอย่างที่ j ในชุดตัวอย่างที่ i เมื่อ $j = 1, 2, \dots, n$ $i = 1, 2, \dots, k$

n = ขนาดตัวอย่าง $\bar{y}_1$ = ค่าเฉลี่ยตัวอย่างในกลุ่มที่ 1

k = ช่วงการสุ่ม $\bar{y}_2$ = ค่าเฉลี่ยตัวอย่างในกลุ่มที่ 2

r = ส่วนที่เหลือจาก $\frac{N}{n}$

3. การเลือกตัวอย่างแบบมีระบบชนิด Modified Balanced Circular (MBC) (Ching-Ho and Fei-Fei, 2006)

ถ้า $x_1, x_2, \dots, x_n$ เป็นหน่วยตัวอย่างขนาด n หน่วยที่สุ่มได้ และทำการจัดเรียงลำดับค่าจากน้อยไปหามากแล้ว y_i คือ ค่าของหน่วยลำดับที่ i และมีความสัมพันธ์เชิงเส้นกับ x_i ดังสมการ $y_i = \alpha + \beta(x_i)$ โดยที่ α, β คือ ค่าคงที่ เมื่อทำการเลือกตัวอย่างด้วยวิธี MBC สำหรับ N เป็นจำนวนคี่ หรือ n เป็นจำนวนคู่ จะให้ค่าเฉลี่ยตัวอย่างเท่ากับค่าเฉลี่ยประชากรด้วยการหาค่าเฉลี่ยตัวอย่างวิธีปกติ $\bar{y}$

เมื่อ $\bar{y} = \sum_{i=1}^n \frac{y_i}{n}$ หรือสามารถหาค่าเฉลี่ยตัวอย่างให้เท่ากับค่าเฉลี่ยประชากรได้ด้วยวิธีการหาค่าเฉลี่ยแบบ Corrected sample mean ($\bar{y}_{cl}$) ดังนี้

$$\bar{y}_{cl} = \frac{1}{n} \left\{ S_{11}y_{x1} + \sum_{i=2}^{n-1} y_{x_i} + S_{12}y_{x_n} \right\}$$

เมื่อ y_{xi} คือ ค่าของหน่วยลำดับที่ x_i ; $i = 1, 2, \dots, n$

$$S_{11} = (E_1 - 2x_n) / (x_1 - x_n)$$

$$S_{12} = (2x_1 - E_1) / (x_1 - x_n)$$

$$E_1 = n(N+1) / 2 - \sum_{i=2}^{n-1} x_i$$

วัตถุประสงค์ของการวิจัย

1. เพื่อศึกษาวิธีการเลือกตัวอย่างแบบมีระบบเชิงเส้น (LSY) , แบบมีระบบเชิงเส้นคู่ (BSY) , แบบมีระบบชนิด Modified Balanced Circular (MBC) และ แบบมีระบบชนิด Mixed Random (MRSS)

2. เพื่อเปรียบเทียบประสิทธิภาพตัวประมาณค่าเฉลี่ยที่ได้จากการเลือกตัวอย่างแบบมีระบบทั้ง 4 เมื่อประชากรมีแนวโน้มเชิงเส้น (Linear Trend)

วิธีดำเนินการวิจัย

งานวิจัยนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของค่าเฉลี่ยตัวอย่างที่ได้จากการเลือกตัวอย่างแบบมีระบบ 4 วิธี คือ การเลือกตัวอย่างแบบมีระบบเชิงเส้น (LSY) , การเลือกตัวอย่างแบบมีระบบเชิงเส้นคู่ (BSY) , การเลือกตัวอย่างแบบมีระบบชนิด Modified Balanced Circular (MBC) และ การเลือกตัวอย่างแบบมีระบบชนิด Mixed Random (MRSS) ศึกษาโดยจำลองข้อมูลจากโปรแกรม SAS 9.1 ซึ่งกำหนดให้ประชากรมีแนวโน้มเชิงเส้น $y_i = 2 + 5i + e_i$ โดยที่ $E(e_i) = 0$, $Var(e_i) = i^2$ เมื่อ $g = 0, 1$, 2 โดยกำหนดขนาดประชากรและตัวอย่างดังนี้

1. $N = 1,500$	$n = 277$	7. $N = 500$	$n = 101$
2. $N = 1,500$	$n = 198$	8. $N = 500$	$n = 65$
3. $N = 1,000$	$n = 170$	9. $N = 100$	$n = 13$
4. $N = 1,000$	$n = 108$	10. $N = 100$	$n = 8$
5. $N = 700$	$n = 124$	11. $N = 25$	$n = 10$
6. $N = 700$	$n = 90$	12. $N = 25$	$n = 7$

1. ดำเนินการเลือกหน่วยตัวอย่างจากประชากรด้วยวิธีการเลือกตัวอย่างแบบมีระบบทั้ง 4 วิธี คำนวณค่าเฉลี่ยตัวอย่างและค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (MSE) ดังนี้

$$MSE(\bar{y}) = \frac{\sum_{h=1}^{2,000} (\bar{y}_h - \mu)^2}{2,000}$$

โดยที่ $\bar{y}_h$ คือ ค่าเฉลี่ยตัวอย่างในการทดลองซ้ำครั้งที่ h
 μ คือ ค่าเฉลี่ยประชากร

2. เปรียบเทียบประสิทธิภาพของการเลือกตัวอย่างแบบมีระบบ 4 วิธี ด้วยค่าประสิทธิภาพเชิงสัมพัทธ์ (Relative Efficiency : RE) ดังนี้

$$RE = \frac{MSE(\hat{\theta}_1)}{MSE(\hat{\theta}_2)}$$

กำหนดให้ $\hat{\theta}_1$ และ $\hat{\theta}_2$ เป็นตัวประมาณค่าของพารามิเตอร์ θ

สรุปผลการวิจัย

จากการเปรียบเทียบค่าเฉลี่ยตัวอย่างด้วยวิธีการเลือกตัวอย่างแบบมีระบบทั้ง 4 วิธี แสดงผลการศึกษาดังตารางที่ 1 และ 2

ตารางที่ 1 ค่าประสิทธิภาพเชิงสัมพัทธ์ที่ได้จากการเลือกตัวอย่างแบบ MRSS เปรียบเทียบกับการเลือกตัวอย่างแบบ MBC , BSY และ LSY

N	n	r	g	$\frac{MBC}{MRSS}$	$\frac{BSY}{MRSS}$	$\frac{LSY}{MRSS}$
1500	277	115	0	0	0.002	0.004
			1	0	0.002	0.004
			2	0	0.160	0.157
	198	114	0	0.034	0.006	0.012
			1	0.035	0.007	0.012
			2	0.324	0.286	0.281
1000	170	150	0	0.239	0.007	0.076
			1	0.228	0.085	0.078
			2	0.746	0.743	0.682
	108	28	0	0.041	0.167	0.014
			1	0.042	0.008	0.014
			2	0.228	0.213	0.201
700	124	80	0	0.068	0.018	0.022
			1	0.044	0.012	0.016
			2	0.325	0.336	0.315
	90	70	0	0	0.174	0.203
			1	0	0.176	0.211
			2	0	0.801	0.760

ตารางที่ 1 (ต่อ)

N	n	r	g	$\frac{MBC}{MRSS}$	$\frac{BSY}{MRSS}$	$\frac{LSY}{MRSS}$
500	101	96	0	0	0.621	0.466
			1	0	0.627	0.455
			2	0	0.956	0.904
	65	45	0	0	0.059	0.080
			1	0	0.067	0.084
			2	0	0.493	0.522
100	13	9	0	0	0.271	0.389
			1	0	0.265	0.377
			2	0	0.684	0.687
	8	4	0	0.926	0.156	0.299
			1	2.095	0.160	0.295
			2	0.980	0.411	0.507
25	10	5	0	1.053	0.519	0.299
			1	0.980	0.492	0.285
			2	0.666	0.865	0.520
	7	4	0	0.003	0.277	0.429
			1	0.035	0.312	0.441
			2	0.374	0.610	0.564

ตารางที่ 2 ค่าประสิทธิภาพเชิงสัมพัทธ์ที่ได้จากการเปรียบเทียบกันระหว่างทางเลือกตัวอย่าง
แบบ MBC BSY และ LSY

N	n	r	g	$\frac{MBC}{LSY}$	$\frac{BSY}{LSY}$	$\frac{MBC}{BSY}$
1500	277	115	0	0	0.521	0
			1	0	0.543	0
			2	0	1.018	0

ตารางที่ 2 (ต่อ)

N	n	r	g	$\frac{MBC}{LSY}$	$\frac{BSY}{LSY}$	$\frac{MBC}{BSY}$
1000	198	114	0	2.869	0.562	5.103
			1	2.984	0.636	4.690
			2	1.155	1.021	1.132
	170	150	0	3.164	1.202	2.632
			1	2.942	1.106	2.659
			2	1.094	1.089	1.005
	108	28	0	3.029	0.522	5.799
			1	2.915	0.531	5.486
			2	1.133	1.059	1.070
700	124	80	0	3.046	0.799	3.808
			1	2.853	0.764	3.735
			2	1.033	1.068	0.967
	90	70	0	0	0.859	0
			1	0	0.834	0
			2	0	1.054	0
	101	96	0	0	1.334	0
			1	0	1.378	0
			2	0	1.058	0
500	65	45	0	0	0.742	0
			1	0	0.802	0
			2	0	0.944	0
	13	9	0	0	0.698	0
			1	0	0.704	0
			2	0	0.995	0
	8	4	0	3.099	0.521	5.952
			1	7.095	0.543	13.065
			2	1.934	0.811	2.385

ตารางที่ 2 (ต่อ)

N	n	r	g	$\frac{MBC}{LSY}$	$\frac{BSY}{LSY}$	$\frac{MBC}{BSY}$
25	10	5	0	3.517	1.733	2.029
			1	3.437	1.727	1.990
			2	1.280	1.664	0.769
	7	4	0	0.006	0.645	0.009
			1	0.079	0.708	0.111
			2	0.664	1.081	0.614

จากตารางที่ 1 และ ตารางที่ 2 แสดงให้เห็นว่า

1. ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (MSE) ของการเลือกตัวอย่างด้วยวิธี MRSS มีแนวโน้ม สูงกว่า การเลือกตัวอย่างด้วยวิธี MBC , BSY และ LSY
2. การเลือกตัวอย่างด้วยวิธี MBC มีค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (MSE) สูงกว่า การเลือกตัวอย่างด้วยวิธี LSY และ BSY แต่เมื่อ N เป็นจำนวนคี่ หรือ n เป็นจำนวนคู่ การเลือกตัวอย่างด้วยวิธี MBC จะมีค่า MSE ที่ต่ำกว่าและเท่ากับ 0
3. โดยรวมแล้วการเลือกตัวอย่างด้วยวิธี BSY มีค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (MSE) ใกล้เคียงกับ การเลือกตัวอย่างด้วยวิธี LSY

อภิปรายผล

จากการศึกษาเปรียบเทียบประสิทธิภาพค่าเฉลี่ยตัวอย่างที่ได้จากการเลือกตัวอย่างแบบมีระบบ ทั้ง 4 วิธี ผลการศึกษาพบว่าค่าเฉลี่ยตัวอย่างที่ได้จากการเลือกตัวอย่างด้วยวิธี LSY และ BSY จะให้ค่า MSE ต่ำสุด แต่เมื่อขนาดประชากรเป็นจำนวนคี่ หรือ ขนาดตัวอย่างเป็นจำนวนคู่ การเลือกตัวอย่างด้วยวิธี MBC ให้ค่า MSE ต่ำที่สุด และเท่ากับ 0

ข้อเสนอแนะ

1. ข้อเสนอแนะจากผลการวิจัยครั้งนี้

1.1 การเลือกตัวอย่างจากประชากรที่มีแนวโน้มเชิงเส้น ถ้าขนาดประชากร (N) เป็นจำนวนคี่ หรือขนาดตัวอย่าง (n) เป็นจำนวนคู่ ควรเลือกใช้การเลือกตัวอย่างวิธี MBC เนื่องจากให้ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (MSE) = 0

1.2 การเลือกตัวอย่างจากประชากรที่มีแนวโน้มเป็นเชิงเส้น ถ้าขนาดประชากร (N) เป็นจำนวนคู่ และ ขนาดตัวอย่าง (n) เป็นจำนวนคี่ ควรเลือกใช้การเลือกตัวอย่างด้วยวิธี LSY หรือ BSY

2. ข้อเสนอแนะในการทำวิจัยครั้งต่อไป

2.1 ควรทำการเปรียบเทียบประสิทธิภาพค่าเฉลี่ยตัวอย่างที่ได้จากการเลือกตัวอย่างเมื่อประชากรมีแนวโน้มเป็นลักษณะอื่น

กิตติกรรมประกาศ

การศึกษาค้นคว้าวิทยานิพนธ์ฉบับนี้ได้สำเร็จด้วยความอนุเคราะห์จาก อาจารย์อำไพ ทองธีรภาพ อาจารย์ที่ปรึกษาวิทยานิพนธ์ ผู้ซึ่งให้คำปรึกษาแนะนำ ตรวจสอบแก้ไขข้อบกพร่องต่างๆ ด้วยความเอาใจใส่ในทุกขั้นตอนและชี้แนะงานวิจัยมาโดยตลอดและใคร่ขอกราบพระคุณคณาจารย์ภาควิชาสถิติ มหาวิทยาลัย เกษตรศาสตร์ทุกท่านที่ได้ให้ความรู้อย่างดีมาโดยตลอด ผู้วิจัยขอขอบพระคุณด้วยความซาบซึ้งและสำนึกในพระคุณเป็นอย่างสูง

เอกสารอ้างอิง

ภาษาไทย

จามรี ทุกข์สุณ. 2543. การศึกษาเปรียบเทียบประสิทธิภาพของตัวสถิติจากการสุ่มตัวอย่างแบบมีระบบหลายชนิด. วิทยานิพนธ์ปริญญาโท, มหาวิทยาลัยเชียงใหม่

ชรินทร์ คุณุสมุทร และ อำนวย มณีศรีวงศ์กุล. 2540. การศึกษาเปรียบเทียบการสุ่มตัวอย่างแบบมีระบบ 3 วิธี : การสุ่มตัวอย่างแบบมีระบบชนิดวงกลม การสุ่มตัวอย่างแบบมีระบบชนิดใหม่ และการสุ่มตัวอย่างแบบมีระบบชนิดสองเชิงเส้น. ว. วิทย มช. 25(2): 134 – 139.

ภาษาต่างประเทศ

Huang, C.K. 2004. Mixed Random Systematic Sampling Designs. **Metrika** 59: 1-11

Leu, C.H. and F.F. Kao. 2006. Modified balanced circular systematic sampling. **Statistics & Probability Letters** 76: 373–383.