การศึกษาเปรียบเทียบการสุ่มตัวอย่างแบบมีระบบ 3 วิธี : การสุ่มตัวอย่างแบบมีระบบชนิดวงกลม การสุ่มตัวอย่างแบบมีระบบชนิดใหม่ และการสุ่มตัวอย่างแบบมีระบบชนิดสองเชิงเส้น

ชรินทร์ คูคู่สมุทร* และ อำนวย มณีศรีวงศ์กูล**

1. บทนำ

ในบรรดาเทคนิคการสุ่มตัวอย่างทั้งหลาย การสุ่มตัวอย่างแบบมีระบบ (Systematic Sampling) ก็เป็นวิธีหนึ่งที่เป็นที่รู้จักและนิยมใช้กันอย่างแพร่หลาย จุดอ่อนของการสุ่มตัวอย่างแบบนี้ คือ กรณีที่ขนาดของประชากร N ไม่เป็นผลคูณของขนาดของตัวอย่าง n กล่าวคือ เราไม่สามารถเขียนอยู่ในรูป $N = nk$ โดยที่ k เป็นจำนวนเต็ม ในทางปฏิบัติสำหรับกรณีนี้คือ เราจะเลือกจำนวนเต็ม k ที่ใกล้เคียงกับ N/n มากที่สุด แต่ตัวอย่างที่ได้จะมีขนาดไม่เท่ากับ n ตามที่ตั้งใจไว้ กล่าวคือ อาจได้ตัวอย่างที่มีขนาดน้อยกว่า n หรือมากกว่า n จุดอ่อนอีกอย่างหนึ่งของการสุ่มตัวอย่างแบบนี้คือ เราไม่สามารถหาตัวประมาณที่ไม่เอนเอียงของความแปรปรวนของค่าเฉลี่ยของตัวอย่างได้

มีนักสถิติหลายท่านพยายามแก้จุดอ่อนดังกล่าว โดยเสนอการสุ่มตัวอย่างแบบมีระบบชนิดวงกลม (Circular Systematic Sampling : CSY) การสุ่มตัวอย่างแบบมีระบบชนิดใหม่ (New Systematic Sampling : NSY) และการสุ่มตัวอย่างแบบมีระบบชนิดสองเชิงเส้น (Bilinear Systematic Sampling : BSY) โดยที่ CSY และ BSY จะแก้จุดอ่อนในกรณี $N = nk$ กล่าวคือ จะทำให้ได้ขนาดตัวอย่างตามที่ตั้งใจไว้ สำหรับ NSY สามารถใช้ได้ทุกกรณีไม่ว่า N จะเขียนอยู่ในรูปผลคูณของ n และ k หรือไม่ก็ตาม แต่จะให้ตัวประมาณที่ไม่เอนเอียงของความแปรปรวนของค่าเฉลี่ยของตัวอย่าง อย่างไรก็ตาม ยังไม่มีผู้นำวิธีการสุ่มตัวอย่างทั้งสามมาเปรียบเทียบกันว่าวิธีใดจะให้ความถูกต้องหรือแม่นยำกว่ากัน จึงเป็นเหตุผลที่จะนำวิธีการสุ่มตัวอย่างทั้งสามเปรียบเทียบกัน ทั้งนี้เพราะถ้าทราบแผนการสุ่มตัวอย่างที่ดีก็จะนำไปสู่การได้ค่าประมาณของตัวแปรที่ดีด้วย

* รองศาสตราจารย์ ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น

** ผู้ช่วยศาสตราจารย์ ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น

2. วัตถุประสงค์ของการศึกษา

เพื่อศึกษาเปรียบเทียบวิธีการสุ่มตัวอย่างทั้งสามว่าวิธีใดจะให้ค่าประมาณของตัวแปรที่สนใจได้ดีกว่ากัน

3. วิธีดำเนินการวิจัย

ประชากรที่ใช้เป็นประชากรธรรมชาติ 2 ประชากรได้แก่ A ผลผลิตข้าวรายจังหวัดปีเพาะปลูก 2535/36 และ B : จำนวนสหกรณ์การเกษตร จำแนกตามจังหวัดปี 2535/36 โดยที่ประชากร A และ B มีขนาด 69 และ 73 ตามลำดับ ในแต่ละประชากรข้างต้น จะใช้วิธีการสุ่มตัวอย่างแบบ CSY NSY และ BSY โดยใช้ขนาดตัวอย่างต่าง ๆ กัน เริ่มที่ n=12 จนถึง 30 และจะคำนวณความคลาดเคลื่อนมาตรฐานของค่าเฉลี่ยของตัวอย่างในแต่ละวิธี จากนั้นก็นำมาเปรียบเทียบกัน สำหรับการเลือกหน่วยตัวอย่างในประชากร และการคำนวณค่าความคลาดเคลื่อนมาตรฐานทำโดยใช้โปรแกรมคอมพิวเตอร์

4. ผลการวิจัย

ค่าความคลาดเคลื่อนมาตรฐาน (และ C.V) ที่ได้จากการสุ่มแต่ละวิธีแสดงในตารางที่ 1 สำหรับตัวอย่างที่สุ่มจากประชากร A และตารางที่ 2 สำหรับตัวอย่างที่สุ่มจากประชากร B

ในตารางที่ 1 จะเห็นว่าความคลาดเคลื่อนมาตรฐานหรือ C.V ของ $\bar{y}_{NSY}$ จะมีค่ามากที่สุดไม่ว่าขนาดของตัวอย่าง n จะเป็นเท่าไร เมื่อ n=12 และ 13 ความคลาดเคลื่อนมาตรฐานของ $\bar{y}_{CSY}$ จะเล็กที่สุด เมื่อ n มีขนาดใหญ่ขึ้น โดยภาพรวมแล้ว ความคลาดเคลื่อนมาตรฐานของ $\bar{y}_{BSY}$ จะน้อยกว่าของ $\bar{y}_{CSY}$ โดยเฉพาะอย่างยิ่งเมื่อเศษส่วนการสุ่มตัวอย่างเกินกว่า 39% ขึ้นไป ความคลาดเคลื่อนมาตรฐานของ $\bar{y}_{BSY}$ จะลดลงอย่างรวดเร็ว ในขณะที่ทั้งของ CSY และ NSY มีขนาดพอ ๆ กัน ในตารางที่ 2 จะเห็นว่าความคลาดเคลื่อนมาตรฐานหรือ C.V ของ $\bar{y}_{NSY}$ จะมีค่ามากที่สุดเช่นกัน ในขณะที่เมื่อ n=12 13 และ 14 ความคลาดเคลื่อนมาตรฐานของ $\bar{y}_{BSY}$ จะน้อยที่สุด แต่ก็ไม่ต่างจากของ CSY มากเท่าไรนัก โดยภาพรวมแล้วจะคล้ายกับประชากร A

5. สรุปผลการวิจัย

ความคลาดเคลื่อนมาตรฐานของค่าเฉลี่ยของตัวอย่างในแต่ละวิธีจากประชากรทั้งสองมีลักษณะคล้ายกัน ความคลาดเคลื่อนมาตรฐาน หรือ C.V ของ $\bar{y}_{NSY}$ จะมีค่ามากที่สุดไม่ว่าขนาดของตัวอย่าง n จะเป็นเท่าไร ในกรณีตัวอย่างมีขนาดเล็ก ความคลาดเคลื่อนมาตรฐานจะมีค่ามากในทุก

วิธี แต่วิธี NSY จะมีค่ามากเมื่อเทียบกับวิธี CSY และ BSY เมื่อตัวอย่างมีขนาดใหญ่ โดยภาพรวมแล้วความคลาดเคลื่อนมาตรฐานของ $\bar{y}_{BSY}$ จะมีค่าน้อยที่สุด โดยเฉพาะอย่างยิ่ง เมื่อเศษส่วนการสุ่มตัวอย่างเกินกว่า 39% ขึ้นไป ความคลาดเคลื่อนมาตรฐานของ $\bar{y}_{BSY}$ จะลดลงอย่างรวดเร็ว ในขณะที่ความคลาดเคลื่อนมาตรฐานของ $\bar{y}_{CSY}$ และ $\bar{y}_{NSY}$ จะมีขนาดพอ ๆ กัน

6. ข้อเสนอแนะ

เนื่องจากเทคนิคการสุ่มตัวอย่างทั้งสามจะทำให้ได้ขนาดของตัวอย่าง n ตามที่ต้องการ เมื่อ $N \neq nk$ และจากผลการวิจัย ผู้วิจัยมีข้อเสนอแนะ ดังนี้

1. เมื่อประชากรที่ศึกษาเป็นประชากรธรรมชาติ ถ้าตัวอย่างมีขนาดเล็กอาจเลือกใช้วิธี CSY หรือ BSY ถ้าตัวอย่างที่ใช้มีขนาดใหญ่ควรใช้วิธี BSY

2. ศึกษาเชิงทฤษฎีเกี่ยวกับความคลาดเคลื่อนมาตรฐานของค่าเฉลี่ยของตัวอย่างจากวิธีการสุ่มตัวอย่างทั้งสาม อีกทั้งควรทำวิจัยเพิ่มเติมกับประชากรที่มีลักษณะอื่น ๆ เช่น linear trend priodic variation หรือ auto-correlated

7. กิตติกรรมประกาศ

ขอขอบคุณคณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น ที่ได้ให้ทุนการวิจัยครั้งนี้ ขอขอบคุณภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น ที่ให้ความสะดวกในการทำวิจัย และท้ายนี้ขอขอบคุณผู้ให้ความช่วยเหลือที่ไม่ได้กล่าวนามไว้ ณ ที่นี้ ซึ่งมีส่วนทำให้งานวิจัยนี้สำเร็จลุล่วงด้วยดี

ตารางที่ 1 ความคลาดเคลื่อนมาตรฐานของ $\bar{y}_{CSY}$ $\bar{y}_{NSY}$ และ $\bar{y}_{BSY}$ เมื่อ N = 69

N	$\sqrt{V(\bar{y}_{CSY})}$	(C.V)	$\sqrt{V(\bar{y}_{NSY})}$	(C.V)	$\sqrt{V(\bar{y}_{BSY})}$	(C.V)
12	29,724.9080*	(11.86)	77,096.1244	(30.77)	40,452.1215	(16.14)
13	41,143.7615*	(16.42)	67,897.9836	(27.10)	43,377.7449	(17.31)
14	38,355.9236	(15.31)	59,780.3074	(23.86)	25,390.3246*	(10.13)
15	38,845.0922	(15.50)	50,277.4445	(20.07)	22,589.9345*	(9.02)
16	25,852.7200	(10.32)	63,385.9538	(25.30)	25,138.2007*	(10.03)
17	22,290.5268*	(8.90)	57,268.3753	(22.86)	25,126.7252	(10.03)
18	22,565.6570	(9.01)	50,728.0365	(20.25)	17,210.4203*	(6.87)
19	25,123.5051*	(10.03)	44,428.8512	(17.73)	25,902.6732	(10.34)
20	32,133.7084	(12.82)	37,567.2902	(14.99)	18,223.9272*	(7.27)
21	26,069.8316	(10.42)	31,251.7014	(12.47)	21,361.0635*	(8.53)
22	21,637.9982	(8.64)	25,795.5717	(10.30)	17,192.2418*	(6.86)
23	19,168.3437*	(7.65)	59,001.0771	(23.55)	19,168.3437*	(7.56)
24	21,262.7973*	(8.49)	54,555.1556	(21.77)	30,642.4501	(12.23)
25	24,263.4524*	(9.68)	50,111.9843	(20.00)	29,842.6157	(11.91)
26	27,586.3222	(11.01)	45,748.9784	(18.26)	23,442.2993*	(9.36)
27	30,561.5183	(12.24)	40,984.0412	(16.36)	5,679.4451*	(2.27)
28	35,417.2904	(14.14)	36,295.1915	(14.49)	8,443.3153*	(3.37)
29	30,644.0064	(12.23)	31,398.7598	(12.53)	5,717.4147*	(2.28)
30	25,639.8660	(10.35)	26,443.5008	(10.55)	3,306.6204*	(1.32)

* ความคลาดเคลื่อนมาตรฐานที่น้อยที่สุด

ตารางที่ 2 ความคลาดเคลื่อนมาตรฐานของ $\bar{y}_{CSY}$ $\bar{y}_{NSY}$ และ $\bar{y}_{BSY}$ เมื่อ N = 73

N	$\sqrt{V(\bar{y}_{CSY})}$	(C.V)	$\sqrt{V(\bar{y}_{NSY})}$	(C.V)	$\sqrt{V(\bar{y}_{BSY})}$	(C.V)
12	2.4110	(9.79)	4.7278	(19.20)	2.0642*	(8.38)
13	2.4007	(9.75)	4.3877	(17.82)	2.2661*	(9.20)
14	2.1970	(8.92)	3.8062	(15.46)	1.0907*	(4.43)
15	1.9940*	(9.10)	3.2973	(13.39)	2.5728	(10.45)
16	1.9305*	(7.84)	3.9391	(16.00)	2.8514	(11.58)
17	1.4159*	(5.75)	3.5411	(14.38)	1.8850	(7.66)
18	1.1765*	(4.78)	3.1894	(12.95)	1.2627	(5.13)
19	1.1651*	(4.73)	2.8124	(11.42)	1.4894	(6.05)
20	1.2583*	(5.11)	2.4657	(10.02)	1.9868	(8.07)
21	1.9532	(7.93)	2.1314	(8.66)	1.8677*	(7.59)
22	1.7075	(6.94)	1.8598	(7.55)	1.3254*	(5.38)
23	1.5223	(6.18)	3.6109	(14.67)	0.9842*	(4.00)
24	1.4168	(5.75)	3.3423	(13.58)	0.5566*	(2.26)
25	1.4193	(5.76)	3.0910	(12.55)	1.2899*	(5.24)
26	1.4760*	(6.00)	2.8288	(11.49)	1.5254	(6.20)
27	1.5690*	(6.37)	2.5849	(10.50)	1.5958	(6.48)
28	1.6490	(6.70)	2.3329	(9.48)	1.2047*	(4.89)
29	1.7273	(7.02)	2.0718	(8.42)	0.8884*	(3.61)
30	1.7788	(7.23)	1.8252	(7.41)	0.4820*	(1.96)

* ความคลาดเคลื่อนมาตรฐานที่น้อยที่สุด

บรรณานุกรม

ชรินทร์ ดูดู่สมุทร **เทคนิคการสุ่มตัวอย่าง** ขอนแก่น : (ม.ป.ท) 2532

สำนักงานเศรษฐกิจการเกษตร ศูนย์เศรษฐกิจการเกษตร **สถิติการเกษตรของประเทศไทย ปีเพาะปลูก 2535/36** กรุงเทพฯ : ธนโชติการพิมพ์, 2536

Agrawal, M.C., Bersales, Lisa Grace S. and Acusta, Andre P. "Bilinear Systematic Sampling" , UNDP-UP Statistical Education Support Project, Research Conference December 4-6,: 1989 (Mimeographed)

Cochran, W.G. **Sampling Techniques** 3rd. ed. New York: John Wiley & Sons, 1977.

Singh, D and Chaudhary, F.S. **Theory and Analysis of Sample Survey Designs** New Delhi: Wiley Eastrn, 1986.

Singh, D and Singh, P. "New Systematic Sampling" **Journal of Statistical Planning and Inference** 1(1977): 163-177.