บทความวิจัย

ตัวประมาณแบบอัตราส่วนรวมของค่าเฉลี่ยประชากรในการชักตัวอย่างแบบแบ่งเป็นชั้นภูมิ

Combined Ratio Estimator of a Population Mean in Stratified Random Sampling

พฤทธิพัทธ์ โสประดิษฐ์[1*] และอำนวย มณีศรีวงศ์กูล[2]

Phruttipat Sopradit[1*] and Amnuay Maneesriwongul[2]

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อเสนอตัวประมาณแบบอัตราส่วนรวมตัวใหม่ของค่าเฉลี่ยประชากรในการชักตัวอย่างแบบแบ่งเป็นชั้นภูมิ ตัวประมาณที่เสนอนี้ผู้วิจัยประยุกต์จากตัวประมาณของมะลิวรรณ ระฆังทอง [1] ประสิทธิภาพของตัวประมาณที่เสนอเปรียบเทียบกับตัวประมาณแบบอัตราส่วนรวมดั้งเดิมและตัวประมาณแบบอัตราส่วนรวมของ Kadilar และ Cingi [2] โดยใช้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยเชิงทฤษฎี ผลการวิจัยพบว่า ตัวประมาณที่เสนอมีประสิทธิภาพดีกว่าตัวประมาณแบบอัตราส่วนรวมดั้งเดิมและตัวประมาณแบบอัตราส่วนรวมของ Kadilar และ Cingi (2005) [2] ในทุกเงื่อนไข ผลเชิงทฤษฎีสอดคล้องกับตัวอย่างการคำนวณเชิงตัวเลข

คำสำคัญ : ตัวประมาณแบบอัตราส่วนรวม ค่าเฉลี่ยประชากร การชักตัวอย่างแบบแบ่งเป็นชั้นภูมิ

Abstract

The objective of this research is to propose a new combined ratio estimator of a population mean in stratified random sampling. This estimator is applied from the estimator proposed by Rakungthong M. [2] The efficiency of proposed estimator is compared with the traditional combined ratio estimator and the combined ratio estimator proposed by Kadilar and Cingi [2] on the basis of theoretical mean square error. The research result is found that the proposed estimator is more efficient than both the traditional combined ratio estimator and Kadilar and Cingi [2] combined ratio estimator in all conditions. The theoretical results are supported by a numeric example.

Keywords : Combined ratio estimator, Population mean, Stratified random sampling

[1] นักศึกษาบัณฑิตศึกษา ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น ขอนแก่น 40002

[2] รองศาสตราจารย์ ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น ขอนแก่น 40002

[*] Corresponding author : E-mail: hadezwarrior@hotmail.com โทรศัพท์: 090-6426848

จากงานประชุมวิชาการระดับชาติ มหาวิทยาลัยทักษิณ ครั้งที่ 22 ปี 2555

บทนำ

การชักตัวอย่างแบบแบ่งประชากรออกเป็นชั้นภูมิเป็นวิธีที่นักวิจัยเชิงสำรวจนิยมใช้กันมากวิธีหนึ่ง การชักตัวอย่างวิธีนี้จะแบ่งประชากรขนาด N ออกเป็นประชากรย่อย ที่เรียกว่าชั้นภูมิ โดยแต่ละชั้นภูมิไม่ซ้อนกัน และมีขนาดเป็น $N_1, N_2, \ldots, N_L$ โดยที่ $\sum_{h=1}^{L} N_h = N$ จากนั้นชักตัวอย่างขนาด $n_1, n_2, \ldots, n_L$ ในแต่ละชั้นภูมิด้วยวิธีการชักตัวอย่างแบบอย่างง่าย โดยที่ $\sum_{h=1}^{L} n_h = n$

ให้ y เป็นตัวแปรที่สนใจศึกษา (Study Variable) และ x เป็นตัวแปรช่วย (Auxiliary Variable)ที่ทราบค่าตัวประมาณแบบอัตราส่วนรวมดั้งเดิม (Traditional Combined Ratio Estimator) ของค่าเฉลี่ยประชากร ($\bar{Y}$) คือ

$$\bar{y}_{CR} = \frac{\bar{y}_{st}}{\bar{x}_{st}} \bar{X} \tag{1}$$

โดยที่ $\bar{y}_{st} = \sum_{h=1}^{L} W_h \bar{y}_h$ และ $\bar{x}_{st} = \sum_{h=1}^{L} W_h \bar{x}_h$ เป็นค่าเฉลี่ยของตัวอย่างที่ชักตัวอย่างแบบแบ่งประชากรออกเป็นชั้นภูมิ และ $W_h = \frac{N_h}{N}$ คือน้ำหนักของชั้นภูมิที่ h และมีความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Square Error: MSE)

$$MSE(\bar{y}_{CR}) = \sum_{h=1}^{L} W_h^2 \theta_h (S_{yh}^2 + R^2 S_{xh}^2 - 2R\rho_h S_{yh} S_{xh}) \tag{2}$$

โดยที่ $\theta_h = (\frac{1}{n_h} - \frac{1}{N_h})$

$R = \frac{\bar{Y}}{\bar{X}}$

S_{yh}^2 คือความแปรปรวนของตัวแปร y ในชั้นภูมิที่ h

S_{xh}^2 คือความแปรปรวนของตัวแปร x ในชั้นภูมิที่ h

และ ρ_h คือสัมประสิทธิ์สหสัมพันธ์ของตัวแปร y และ x ในชั้นภูมิที่ h

มะลิวรรณ ระฆังทอง (2552) [1] เสนอตัวประมาณแบบอัตราส่วนของค่าเฉลี่ยประชากรในการชักตัวอย่างแบบอย่างง่าย

$$\bar{y}_{RM} = \frac{\bar{y} + b(\bar{X} - \bar{x})}{\bar{x}\rho + X} (\bar{X}\rho + X) \tag{3}$$

ที่มี

$$MSE(\bar{y}_{RM}) = \frac{1-f}{n} \left(R_{RM}^2 S_x^2 + S_y^2 (1 - \rho^2) \right) \tag{4}$$

โดยที่ $f = \frac{n}{N}$

$R_{RM} = \frac{\bar{Y}\rho}{\bar{X}\rho + X}$

S_x^2 คือความแปรปรวนของตัวแปร x

S_y^2 คือความแปรปรวนของตัวแปร y

และ ρ คือสัมประสิทธิ์สหสัมพันธ์ของตัวแปร y และ x

พบว่าตัวประมาณนี้มีประสิทธิภาพดีกว่าตัวประมาณของ Kadilar และ Cingi ในทุกเงื่อนไข [4]

Kadilar และ Cingi [2] เสนอตัวประมาณแบบอัตราส่วนรวม

$$\bar{y}_{KC} = k^* \bar{y}_{CR} \tag{5}$$

โดย k^* เป็นค่าคงที่ที่ทำให้ $\bar{y}_{KC}$ มีความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุด และมีค่าดังนี้

$$MSE(\bar{y}_{KC}) = (k^* - 1)^2 \bar{Y}^2 + k^{*2} MSE(\bar{y}_{CR}) \tag{6}$$

โดยที่

$$k^* = \frac{\bar{Y}^2}{\bar{Y}^2 + MSE(\bar{y}_{CR})} \tag{7}$$

เมื่อ แทนค่า k^* ลงใน (6) จะได้

$$MSE(\bar{y}_{KC})_{opt} = \frac{\bar{Y}^2 MSE(\bar{y}_{\mathrm{CR}})}{\bar{Y}^2 + MSE(\bar{y}_{\mathrm{CR}})} \tag{8}$$

พบว่า ตัวประมาณ $\overline{y}_{\mathrm{KC}}$ มีประสิทธิภาพดีกว่าตัวประมาณ $\overline{y}_{\mathrm{CR}}$ ในทุกเงื่อนไข

ในการวิจัยครั้งนี้ ผู้วิจัยจะเสนอตัวประมาณแบบอัตราส่วนรวมในการชักตัวอย่างแบบแบ่งประชากรออกเป็นชั้นภูมิ โดยพัฒนาจากตัวประมาณของมะลิวรรณ ระฆังทอง [1]

วิธีการวิจัย

มีวิธีการวิจัยดำเนินการเป็นขั้นตอน ดังนี้

1. ผู้วิจัยเสนอตัวประมาณแบบอัตราส่วนรวมของค่าเฉลี่ยประชากรในการชักตัวอย่างแบบแบ่งเป็นชั้นภูมิ

$$\bar{y}_P = \frac{\bar{y}_{st} + b(\bar{X} - \bar{x}_{st})}{\bar{x}_{st}\rho_c + X}(\bar{X}\rho_c + X) \tag{9}$$

โดยที่ $\rho_c = \dfrac{\sum_{h=1}^{L} W_h^2 \theta_h \rho_h S_{yh} S_{xh}}{\sqrt{\sum_{h=1}^{L} W_h^2 \theta_h S_{yh}^2}\sqrt{\sum_{h=1}^{L} W_h^2 \theta_h S_{xh}^2}}$

2. หาความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณที่นำเสนอ โดยใช้อนุกรมเทย์เลอร์อันดับที่ 1

$$h(\bar{x}_{st}, \bar{y}_{st}) \cong h(\bar{X}, \bar{Y}) + \left.\frac{\partial h(\bar{x}_{st}, \bar{y}_{st})}{\partial \bar{x}_{st}}\right|_{\bar{X},\bar{Y}} (\bar{x}_{st} - \bar{X}) + \left.\frac{\partial h(\bar{x}_{st}, \bar{y}_{st})}{\partial \bar{y}_{st}}\right|_{\bar{X},\bar{Y}} (\bar{y}_{st} - \bar{Y})$$

ในที่นี้ $\bar{y}_P = h(\bar{x}_{st}, \bar{y}_{st})$ และ $\bar{Y} = h(\bar{X}, \bar{Y})$

3. เปรียบเทียบประสิทธิภาพของตัวประมาณแบบอัตราส่วนรวมที่เสนอกับตัวประมาณแบบอัตราส่วนรวมดั้งเดิมและตัวประมาณของ Kadilar และ Cingi [2] โดยใช้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย โดยที่ตัวประมาณ $\hat{\theta}_1$ เป็นตัวประมาณที่มีประสิทธิภาพดีกว่า $\hat{\theta}_2$ ถ้า MSE ($\hat{\theta}_1$)<MSE($\hat{\theta}_2$)

จากงานประชุมวิชาการระดับชาติ มหาวิทยาลัยทักษิณ ครั้งที่ 22 ปี 2555

ผลการวิจัย

1. ความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณ หาได้จากอนุกรมเทย์เลอร์อันดับที่ 1

$$\mathrm{h}(\bar{x}_{st},\bar{y}_{st}) \cong h(\bar{X},\bar{Y}) + \left.\frac{\partial h(\bar{x}_{st},\bar{y}_{st})}{\partial \bar{x}_{st}}\right|_{\bar{X},\bar{Y}} (\bar{x}_{st}-\bar{X}) + \left.\frac{\partial h(\bar{x}_{st},\bar{y}_{st})}{\partial \bar{y}_{st}}\right|_{\bar{X},\bar{Y}} (\bar{y}_{st}-\bar{Y})$$

$$\bar{y}_P - \bar{Y} \cong \left.\frac{\partial}{\partial \bar{x}_{st}}\left(\frac{\bar{y}_{st}+b(\bar{X}-\bar{x}_{st})}{\bar{x}_{st}\rho_c + X}(\bar{X}\rho_c + X)\right)\right|_{\bar{X},\bar{Y}} (\bar{x}_{st}-\bar{X}) + \left.\frac{\partial}{\partial \bar{y}_{st}}\left(\frac{\bar{y}_{st}+b(\bar{X}-\bar{x}_{st})}{\bar{x}_{st}\rho_c + X}(\bar{X}\rho_c + X)\right)\right|_{\bar{X},\bar{Y}} (\bar{y}_{st}-\bar{Y})$$

$$\bar{y}_P - \bar{Y} \cong \left(-\frac{\bar{Y}\rho_C}{\bar{X}\rho_C + X} - b\right)(\bar{x}_{st}-\bar{X}) + (\bar{y}_{st}-\bar{Y})$$

$$MSE(\bar{y}_P) = E(\bar{y}_P - \bar{Y})^2$$

$$\cong \left(\frac{\bar{Y}^2\rho_C^2}{(\bar{X}\rho_C + X)^2} + \frac{2\beta\bar{Y}\rho_C}{\bar{X}\rho_C + X} + \beta^2\right)V(\bar{x}_{st}) + 2\left(-\frac{\bar{Y}\rho_C}{\bar{X}\rho_C + X} - \beta\right)Cov(\bar{x}_{st},\bar{y}_{st}) + V(\bar{y}_{st})$$

แทน $V(\bar{x}_{st}) = \sum_{h=1}^{L} W_h^2\,\theta_h S_{xh}^2$, $V(\bar{y}_{st}) = \sum_{h=1}^{L} W_h^2\,\theta_h S_{yh}^2$, $Cov(\bar{x}_{st},\bar{y}_{st}) = \sum_{h=1}^{L} W_h^2\,\theta_h\,\rho_h S_{yh} S_{xh}$,

$\beta = \frac{\sum_{h=1}^{L} W_h^2\theta_h\rho_h S_{yh}S_{xh}}{\sum_{h=1}^{L} W_h^2\theta_h S_{xh}^2}$ และให้ $R_P = \frac{\bar{Y}\rho_C}{\bar{X}\rho_C + X}$ จะได้

$$MSE(\bar{y}_P) \cong \sum_{h=1}^{L} W_h^2\theta_h\left[R_P^2 S_{xh}^2 + S_{yh}^2\right] - \frac{\left(\sum_{h=1}^{L} W_h^2\theta_h\rho_h S_{yh}S_{xh}\right)^2}{\sum_{h=1}^{L} W_h^2\theta_h S_{xh}^2} \qquad (10)$$

2. เปรียบเทียบประสิทธิภาพของตัวประมาณเชิงทฤษฎี

2.1 ประสิทธิภาพของตัวประมาณที่เสนอกับตัวประมาณแบบอัตราส่วนรวมดั้งเดิม

ตัวประมาณ yP จะมีประสิทธิภาพดีกว่าตัวประมาณ ($\overline{y}_{CR}$) ถ้า

$$MSE(\bar{y}_P) - MSE(\bar{y}_{CR}) < 0$$

แทนค่า $MSE(\bar{y}_P)$ และ $MSE(\bar{y}_{CR})$ จะได้

$$\sum_{h=1}^{L} W_h^2\theta_h\left[R_P^2 S_{xh}^2 + S_{yh}^2\right] - \frac{\left(\sum_{h=1}^{L} W_h^2\theta_h\rho_h S_{yh}S_{xh}\right)^2}{\sum_{h=1}^{L} W_h^2\theta_h S_{xh}^2} - \sum_{h=1}^{L} W_h^2\theta_h\left(S_{yh}^2 + R^2 S_{xh}^2 - 2R\rho_h S_{yh}S_{xh}\right) < 0$$

ให้ $A = \sum_{h=1}^{L} W_h^2\theta_h S_{xh}^2$, $B = \sum_{h=1}^{L} W_h^2\theta_h S_{yh}^2$ และ $C = \sum_{h=1}^{L} W_h^2\theta_h\rho_h S_{yh}S_{xh}$

จะได้ $R_P^2 A + B - \frac{C^2}{A} - (B + R^2 A - 2RC) < 0$

$$(R_P^2 - R^2)A < \frac{C^2}{A} - 2RC$$

หรือ $R_P^2 - R^2 < \frac{C}{A^2}(C - 2RA)$ (11)

พิจารณา

$$R_P^2 = \left(\frac{\bar{Y}\rho_C}{\bar{X}\rho_C + X}\right)^2 = \left(\frac{\bar{Y}\rho_C}{\bar{X}(\rho_C + N)}\right)^2 = R^2\left(\frac{\rho_C}{\rho_C + N}\right)^2$$

แทน R_P^2 ด้วย $R^2\left(\frac{\rho_C}{\rho_C+N}\right)^2$ ใน (11) จะได้

$$R^2\left(\frac{\rho_C}{\rho_C + N}\right)^2 - R^2 < \frac{C}{A^2}(C - 2RA)$$

$$\left(\left(\frac{\rho_C}{\rho_C + N}\right)^2 - 1\right)R^2 < \frac{C}{A^2}(C - 2RA)$$

เมื่อ N มีขนาดใหญ่ $\left(\frac{\rho_C}{\rho_C+N}\right)^2 \to 0$ จะได้

$$-R^2 < \frac{C}{A^2}(C - 2RA)$$

$$R^2A^2 + C^2 - 2RAC > 0$$

หรือ

$$(RA - C)^2 > 0 \quad (12)$$

ซึ่งเงื่อนไข (12) เป็นจริงเสมอ ดังนั้น ตัวประมาณที่เสนอ $\overline{y}_p$ มีประสิทธิภาพดีกว่าตัวประมาณแบบอัตราส่วนรวมดั้งเดิมในทุกเงื่อนไข

2.2 ประสิทธิภาพของตัวประมาณที่เสนอกับตัวประมาณของ Kadilar และ Cingi (2005)

$$MSE(\bar{y}_P) - MSE(\bar{y}_{KC})_{opt} = MSE(\bar{y}_P) - \frac{\bar{Y}^2 MSE(\bar{y}_{CR})}{\bar{Y}^2 + MSE(\bar{y}_{CR})}$$

$$= \frac{\bar{Y}^2(MSE(\bar{y}_P) - MSE(\bar{y}_{CR})) + MSE(\bar{y}_P)MSE(\bar{y}_{CR})}{(\bar{Y}^2 + MSE(\bar{y}_{CR}))} < 0$$

ดังนั้น ตัวประมาณ $\overline{y}_P$ จะมีประสิทธิภาพดีกว่าตัวประมาณ $\bar{y}_{KC}$ ถ้า $MSE(\bar{y}_P) < MSE(\bar{y}_{CR})$และในทำนองเดียวกับ 2.1 เราสามารถสรุปได้ว่าตัวประมาณ yP จะมีประสิทธิภาพดีกว่าตัวประมาณ yKC ในทุกเงื่อนไข

1.1 ตัวอย่างการคำนวณ

ประชากรที่ใช้เป็นตัวอย่างการคำนวณจะใช้ข้อมูลของ Murthy [3] โดยที่จำนวนโรงงานในพื้นที่เป็นตัวแปร y ที่สนใจศึกษา และจำนวนคนงานเป็นตัวแปรช่วย x จาก 80 โรงงานโดยแบ่งเป็น 4 ชั้นภูมิ ตามลักษณะของตัวแปรช่วย x คือ $x < 100$, $100 \leq x < 200$, $200 \leq x < 500$ และ $x \geq 500$ ลักษณะของประชากรแสดงดังตาราง

จากงานประชุมวิชาการระดับชาติ มหาวิทยาลัยทักษิณ ครั้งที่ 22 ปี 2555

ตารางที่ 1 ลักษณะและค่าพารามิเตอร์ของประชากร

Total	Stratum →	1	2	3	4
$N = 80$	N_h	25	23	16	16
$n = 45$	n_h	14	13	9	9
$\bar{X} = 284.75$	$\bar{X}_h$	71.0	140.696	362.937	749.5
$\bar{Y} = 5182.64$	$\bar{Y}_h$	3156.64	4766.22	6334.19	7795.31
S_x=270.495	S_{xh}	14.6116	28.0364	91.3823	174.463
$S_y = 1835.66$	S_{yh}	740.012	515.697	501.399	653.09
$R_P = 0.151$	ρ_h	0.8167	0.8231	0.9582	0.9805
$R = 18.201$	θ_h	0.03143	0.03344	0.04861	0.04861
$\rho_c = 0.67079$	W_h^2	0.09766	0.08266	0.04	0.04

ตารางที่ 2 MSE ของตัวประมาณแบบอัตราส่วนรวม

ตัวประมาณ	MSE
$\bar{y}_{CR}$	16456.65
$\bar{y}_{KC}$	16446.57
$\bar{y}_P$	**2055.73**

จากตารางที่ 2 จะเห็นว่า MSE ของตัวประมาณแบบอัตราส่วนรวมดั้งเดิมและตัวประมาณของ Kadilar และ Cingi [4] มีค่าใกล้เคียงกัน ในขณะที่ตัวประมาณที่ผู้วิจัยเสนอมี MSE น้อยกว่าอย่างชัดเจน

สรุปผลการวิจัย

งานวิจัยนี้ผู้วิจัยเสนอตัวประมาณแบบอัตราส่วนรวมของค่าเฉลี่ยประชากรในการชักตัวอย่างแบบแบ่งเป็นชั้นภูมิ และเปรียบเทียบประสิทธิภาพของตัวประมาณที่เสนอกับตัวประมาณแบบอัตราส่วนรวมดั้งเดิมและตัวประมาณของ Kadilar และ Cingi [2] ผลการวิจัยพบว่าตัวประมาณที่เสนอมีประสิทธิภาพดีกว่าตัวประมาณแบบอัตราส่วนรวมดั้งเดิมและตัวประมาณของ Kadilar และ Cingi [2] ในทุกเงื่อนไข และตัวอย่างการคำนวณสอดคล้องกับผลเชิงทฤษฎี

เอกสารอ้างอิง

[1] มะลิวรรณ ระฆังทอง. (2552). **ตัวประมาณอัตราส่วนของค่าเฉลี่ยประชากรในการสุ่มตัวอย่าง แบบอย่างง่าย.** วิทยานิพนธ์ วิทยาศาสตรมหาบัณฑิต. ขอนแก่น : มหาวิทยาลัยขอนแก่น.

[2] Kadilar, C. and Cingi H. (2005). "A new ratio estimator in stratified sampling," **Communications in Statistics – Theory and Methods.** 34, 597-602.

[3] Murthy, M.N. (1967). **Sampling Theory and Methods.** India : Statistical Publishing Society.

[4] Kadilar, C. and Cingi H. (2004). "Ratio estimators in simple random sampling," **Applied Mathematics and Computation.** 151, 893-902.

[5] Cochran, W.G. (1977). **Sampling Techniques.** New York : John Wiley and Sons.