

ตัวประมาณแบบผลคูณตัวใหม่ของค่าเฉลี่ยประชากรในการสุ่มตัวอย่างอย่างง่าย A New Product Estimator of a Population Mean in Simple Random Sampling

สุนันทา มงคลสินธุ์¹ และ อำนวย มณีศรีวงศ์กุล^{2*}
Sunanta Mongkonsin¹ and Amnuay Manesriwongul^{2*}

บทคัดย่อ

บทความนี้ผู้วิจัยเสนอตัวประมาณแบบผลคูณตัวใหม่ของค่าเฉลี่ยประชากรในการสุ่มตัวอย่างอย่างง่าย ประสิทธิภาพของตัวประมาณที่เสนอถูกเปรียบเทียบกับตัวประมาณแบบผลคูณสามัญและตัวประมาณแบบผลคูณของ Singh และคณะ [1] โดยใช้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยและค่าร้อยละประสิทธิภาพสัมพัทธ์ผลการวิจัย พบว่า ตัวประมาณที่เสนอมีประสิทธิภาพสูงกว่าตัวประมาณแบบผลคูณสามัญและตัวประมาณแบบผลคูณของ Singh และคณะ [1] แบบมีเงื่อนไข ผลการศึกษาเชิงทฤษฎีของค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสอดคล้องกับการคำนวณเชิงตัวเลข

คำสำคัญ: ค่าเฉลี่ยประชากร ตัวประมาณแบบผลคูณ การสุ่มตัวอย่างอย่างง่าย

Abstract

This paper we propose a new product estimator of a population mean in simple random sampling. The efficiency of this estimator is compared with the conventional product estimator and the product estimator proposed by Singh et al. [1] on the basis of mean square error and percent relative efficiency. It is found that the proposed estimator is more efficient than both the conventional product estimator and the product estimator proposed by Singh et al. [1] under some conditions. In additions, the theoretical of mean square error of the estimators are supported by a numerical illustration.

Keywords : Population Mean, Product Estimator, Simple Random Sampling

¹ นักศึกษาปริญญาโท ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น ขอนแก่น 40002

² รองศาสตราจารย์, ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น ขอนแก่น 40002

* Corresponding author : E-mail : amnuay@kku.ac.th Tel. 043-202222 ext.12258

การสำรวจด้วยตัวอย่างถือเป็นวิธีการทางสถิติที่สำคัญวิธีหนึ่งที่ใช้ในการรวบรวมข้อมูลเนื่องจากในการรวบรวมข้อมูลแต่ละครั้งนั้นอาจเกิดอุปสรรคในการรวบรวมข้อมูล อาทิเช่น ในการรวบรวมข้อมูลอาจมีเวลางบประมาณ หรือทรัพยากรที่จำกัด การสำรวจด้วยตัวอย่างจึงเป็นทางเลือกที่สำคัญที่ใช้ในการรวบรวมข้อมูลเพื่อให้บรรลุต่อวัตถุประสงค์ของงานวิจัยนั้นๆ ภายใต้เงื่อนไขที่แตกต่างกัน แต่สิ่งที่ตรงกันของการสำรวจด้วยตัวอย่างได้แก่การรวบรวมข้อมูลจากตัวอย่างเพื่อวิเคราะห์และประเมินผลถึงลักษณะของประชากรที่สนใจศึกษาตัวอย่างที่ได้มานั้นจึงมีผล ต่อข้อสรุปเกี่ยวกับประชากร ดังนั้นเทคนิคการสำรวจด้วยตัวอย่างจึงมีบทบาทในการเพิ่มประสิทธิภาพของตัวประมาณเพื่อให้ได้มาซึ่งค่าประมาณของพารามิเตอร์ที่ใช้อธิบายลักษณะของประชากร

ในหลายๆ การสำรวจด้วยตัวอย่าง บางครั้งจะพบว่าสารสนเทศของตัวแปรช่วย x (Auxiliary Variable) ที่มีความสัมพันธ์กับตัวแปรที่ต้องการศึกษา y (Study Variable) สามารถนำมาประยุกต์ใช้ในการประมาณค่าได้ กล่าวคือ ถ้าเรามีประชากรขนาด N ซึ่งประกอบไปด้วย (Y_i, X_i) ; $i = 1, 2, 3, \dots, N$ ภายใต้ตัวแปรที่ต้องการศึกษาและทราบค่าตัวแปรช่วย x เราสามารถประมาณค่าเฉลี่ยประชากร $\bar{Y} = \frac{Y}{N}$ โดยที่ $Y = \sum Y_i$ ที่ได้จากการสุ่มตัวอย่างอย่างง่ายแบบไม่ใส่คืน (Simple Random Sampling Without Replacement) ขนาด n จาก N ถ้าความสัมพันธ์ระหว่างตัวแปรที่ต้องการศึกษา y กับตัวแปรช่วย x มีความสัมพันธ์ในทิศทางเดียวกันตัวประมาณดั้งเดิมที่เรามักใช้ คือ $Y_r = \frac{\bar{y}}{\bar{x}} \bar{X}$ ตัวประมาณแบบอัตราส่วน (Ratio Estimator) คือ แต่ถ้าความสัมพันธ์ระหว่างตัวแปรที่ต้องการศึกษา y กับตัวแปรช่วย x มีความสัมพันธ์ในทิศทางตรงข้าม Murthy [2] ได้เสนอให้พิจารณาใช้ตัวประมาณแบบผลคูณ (Product Estimator) คือ

$$\bar{y}_p = \frac{\bar{y} \cdot \bar{x}}{\bar{X}}$$

ที่มีค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Squared Error: MSE)

$$MSE(\bar{y}_p) = \frac{1-f}{n} \bar{Y}^2 [C_y^2 + C_x^2 + 2\rho C_x C_y]$$

เราเรียกตัวประมาณแบบผลคูณข้างต้นว่า ตัวประมาณแบบผลคูณสามัญ (Conventional Product Estimator)

โดยที่

$\bar{y}$ คือ ค่าเฉลี่ยตัวอย่างของตัวแปรที่ต้องการศึกษา y

$\bar{x}$ คือ ค่าเฉลี่ยตัวอย่างของตัวแปรช่วย x

$\bar{X}$ คือ ค่าเฉลี่ยประชากรของตัวแปรช่วย x และทราบค่า

C_x คือ สัมประสิทธิ์ความผันแปร (Coefficient of Variation) ของตัวแปรช่วย x

C_y คือ สัมประสิทธิ์ความผันแปร ของตัวแปรที่ต้องการศึกษา y

ρ คือ สัมประสิทธิ์สหสัมพันธ์ (Correlation Coefficient) ของตัวแปรที่ต้องการศึกษา y กับ x

และ $f = \frac{n}{N}$

จากงานประชุมวิชาการระดับชาติ มหาวิทยาลัยทักษิณ ครั้งที่ 22 ปี 2555

มีนักวิจัยหลายท่านให้ความสนใจในการปรับปรุงตัวประมาณค่าเฉลี่ยประชากรทั้งแบบอัตราส่วนและแบบผลคูณ เพื่อให้มีประสิทธิภาพที่สูงขึ้นกว่าเดิม โดยนำเสนอเสนอแนะของตัวแปรช่วย มาประกอบการประมาณค่าเฉลี่ยประชากร อาทิเช่น Banerjee และ Tiwari [3], Pandey และ Dubey (1988 อ้างถึงใน Khoshnevisan et al., [4]), Singh และคณะ [1], GN Singh [6] และ Upadhyaya, และ Singh [6] เป็นต้น

Khoshnevisan และคณะ [4] พบว่า ตัวประมาณแบบผลคูณที่ปรับปรุงโดยการเพิ่มสารสนเทศของตัวแปรช่วยในกรณีที่ใช้ สัมประสิทธิ์ความโด่ง (Coefficient of Kurtosis:) จะมีประสิทธิภาพสูงกว่าตัวประมาณแบบผลคูณที่เพิ่มสารสนเทศของตัวแปรช่วยตัวอื่นๆ Banerjee และ Tiwari [3] ได้เสนอการใช้สเกลลาร์ลักษณะเฉพาะ (Characterizing Scalar) ในการปรับปรุงตัวประมาณแบบอัตราส่วนของค่าเฉลี่ยประชากร คือ

$$\bar{y}_{ss} = \bar{y} \left[\frac{\bar{X}C_x + 1}{A(\bar{x}C_x + 1) + (1 - A)(\bar{X}C_x + 1)} \right]$$

ที่มีค่าความคลาดเคลื่อนกำลังสองเฉลี่ย

$$MSE(\bar{y}_{ss}) = \gamma_n \bar{Y}^2 [C_y^2 + \alpha C_x^2 (\alpha - 2K)]$$

โดยที่

A คือ สเกลลาร์ลักษณะเฉพาะ

$$\gamma_n = \frac{1 - f}{n}$$

$$\alpha = \frac{\bar{X}C_x}{\bar{X}C_x + 1}$$

$$K = \rho \frac{C_y}{C_x}$$

และ Singh และคณะ (2004) [4] ได้เสนอตัวประมาณแบบผลคูณของค่าเฉลี่ยประชากร คือ

$$\bar{y}_{pr} = \bar{y} \left[\frac{\bar{x} + \beta_2(x)}{\bar{X} + \beta_2(x)} \right]$$

ที่มีค่าความคลาดเคลื่อนกำลังสองเฉลี่ย

$$MSE(\bar{y}_{pr}) = \frac{1-f}{n} \bar{Y}^2 [C_y^2 + \theta_1^2 C_x^2 + 2\theta_1 \rho C_x C_y]$$

โดยที่

$$\theta_1 = \frac{\bar{X}}{\bar{X} + \beta_x(x)}$$

และตัวประมาณ $\bar{y}_{pr}$ จะมีประสิทธิภาพสูงกว่า $\bar{y}_p$ เมื่อ

$$(\theta_1 C_x + \rho C_y)^2 > 0$$

จากงานวิจัยข้างต้นผู้วิจัยจึงสนใจที่จะพัฒนาตัวประมาณแบบผลคูณของค่าเฉลี่ยประชากรตัวใหม่

วิธีการวิจัยและผลการวิจัย

มีขั้นตอนในการดำเนินการ ดังนี้

1. จากแนวคิดของ Banerjee และ Tiwari [3] และ Singh และคณะ [1] ผู้วิจัยนำเสนอตัวประมาณแบบผลคูณใหม่สำหรับประมาณค่าเฉลี่ยประชากร คือ

$$\bar{y}_{pj} = \bar{y} \left[\frac{A(\bar{x} + \beta_2(x)) + (1-A)(\bar{X} + \beta_2(x))}{\bar{X} + \beta_2(x)} \right]$$

2. หาค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณที่เสนอ โดยที่ ความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณ $\hat{\theta}$ ที่ใช้ประมาณค่าพารามิเตอร์ θ คือ $MSE(\hat{\theta}) = E(\hat{\theta} - \theta)^2$

หาค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณ

$$\bar{y}_{pj} = \bar{y} \left[\frac{A(\bar{x} + \beta_2(x)) + (1-A)(\bar{X} + \beta_2(x))}{\bar{X} + \beta_2(x)} \right] \quad (1)$$

ให้ $\bar{y} = \bar{Y} + e_o$ และ $\bar{x} = \bar{X} + e_1$ เนื่องจาก $\bar{y}$ และ $\bar{x}$ เป็นตัวประมาณที่ไม่เอนเอียง (Unbiased Estimator) ของ $\bar{Y}$ และ $\bar{X}$ ตามลำดับ ทำให้ได้

จากงานประชุมวิชาการระดับชาติ มหาวิทยาลัยทักษิณ ครั้งที่ 22 ปี 2555

$$E(e_0) = E(e_1) = 0$$

$$E(e_0^2) = \frac{1-f}{n} \bar{Y}^2 C_y^2$$

$$E(e_1^2) = \frac{1-f}{n} \bar{X}^2 C_x^2$$

และ

$$E(e_0 e_1) = \frac{1-f}{n} \bar{X} \bar{Y} \rho C_x C_y$$

แทนค่า $\bar{y} = \bar{Y} + e_0$ และ $\bar{x} = \bar{X} + e_1$ ลงใน (1) จะได้

$$\bar{y}_{pj} = (\bar{Y} + e_0) \left[\frac{A(\bar{X} + e_1 + \beta_2(x)) + (1-A)(\bar{X} + \beta_2(x))}{\bar{X} + \beta_2(x)} \right]$$

$$= (\bar{Y} + e_0) \left[\frac{Ae_1 + \bar{X} + \beta_2(x)}{\bar{X} + \beta_2(x)} \right]$$

$$= (\bar{Y} + e_0) \left[1 + \frac{Ae_1}{\bar{X} + \beta_2(x)} \right]$$

$$= \bar{Y} + \frac{\bar{Y}Ae_1}{\bar{X} + \beta_2(x)} + e_0 + \frac{Ae_0e_1}{\bar{X} + \beta_2(x)}$$

$$\bar{y}_{pj} - \bar{Y} = e_0 + \frac{Ae_0e_1}{\bar{X} + \beta_2(x)} + \frac{\bar{Y}Ae_1}{\bar{X} + \beta_2(x)} \quad (2)$$

เมื่อพิจารณาอันดับที่ 1 ของการประมาณ (First Order of Approximation) [3] เราสามารถเขียน

$$\bar{y}_{pj} - \bar{Y} = e_0 + \frac{\bar{Y}Ae_1}{\bar{X} + \beta_2(x)}$$

$$MSE(\bar{y}_{pj}) \cong E(\bar{y}_{pj} - \bar{Y})^2$$

$$\cong E \left[\frac{\bar{Y}Ae_1}{\bar{X} + \beta_2(x)} + e_0 \right]^2$$

$$\cong E(e_0^2) + 2A \left[\frac{\bar{Y}}{\bar{X} + \beta_2(x)} \right] E(e_0 e_1) + A^2 \left[\frac{\bar{Y}}{\bar{X} + \beta_2(x)} \right]^2 E(e_1^2)$$

$$\cong \frac{1-f}{n} \bar{Y}^2 C_y^2 + 2 \left[\frac{\bar{Y}A}{\bar{X} + \beta_2(x)} \right] \frac{1-f}{n} \bar{X}\bar{Y}\rho C_x C_y + \left[\frac{\bar{X}A}{\bar{X} + \beta_2(x)} \right]^2 \frac{1-f}{n} \bar{Y}^2 C_x^2$$

กำหนดให้ $\theta = \frac{\bar{X}}{\bar{X} + \beta_2(x)}$ จะได้

$$\begin{aligned} MSE(\bar{y}_{pj}) &\cong \frac{1-f}{n} \bar{Y}^2 [C_y^2 + A^2 \theta^2 C_x^2 + 2A\theta\rho C_x C_y] \\ &\cong \frac{1-f}{n} \bar{Y}^2 [C_y^2 + \theta C_x^2 (A^2 \theta + 2AK)] \end{aligned} \quad (3)$$

โดยที่ $K = \rho \frac{C_y}{C_x}$

ในการหา $MSE(\bar{y}_{pj})$ ที่มีค่าต่ำสุด เราหาอนุพันธ์ใน (3) เทียบกับ A และให้เท่ากับ 0 ทำให้ได้ $A = -\frac{K}{\theta} = A_{opt}$
แทนค่า A_{opt} ลงใน (3) จะได้

$$MSE_{min}(\bar{y}_{pj}) = \frac{1-f}{n} \bar{Y}^2 [1 - \rho^2] C_y^2$$

3. เปรียบเทียบประสิทธิภาพของตัวประมาณแบบผลคูณที่นำเสนอ โดยทำการเปรียบเทียบกับตัวประมาณแบบผลคูณสามัญและตัวประมาณแบบผลคูณของ Singh และคณะ [1] เรากล่าวว่า ตัวประมาณมีประสิทธิภาพสูงกว่าตัวประมาณ $\hat{\theta}_2$ ถ้า $MSE(\hat{\theta}_1) - MSE(\hat{\theta}_2) < 0$ และคำนวณค่าร้อยละประสิทธิภาพสัมพัทธ์ (Percent Relative Efficiency: PRE) ของ $\hat{\theta}_1$ เทียบกับ $\hat{\theta}_2$

โดยที่ $PRE = \frac{MSE(\hat{\theta}_2)}{MSE(\hat{\theta}_1)} \times 100$

3.1 เปรียบเทียบประสิทธิภาพของตัวประมาณ $\bar{y}_{pj}$ กับ $\bar{y}_p$ เจริญฤทธิ์
เรามีเงื่อนไข

$$MSE_{min}(\bar{y}_{pj}) - MSE(\bar{y}_p) < 0$$

$$\frac{1-f}{n} \bar{Y}^2 [1 - \rho^2] C_y^2 - \frac{1-f}{n} \bar{Y}^2 [C_y^2 + C_x^2 + 2\rho C_x C_y] < 0$$

$$\frac{1-f}{n} \bar{Y}^2 [(1 - \rho^2) C_y^2 - (C_y^2 + C_x^2 + 2\rho C_x C_y)] < 0$$

จากงานประชุมวิชาการระดับชาติ มหาวิทยาลัยทักษิณ ครั้งที่ 22 ปี 2555

ทำให้ได้

$$-\left[\rho^2 C_y^2 + C_x^2 + 2\rho C_x C_y\right] < 0$$

ทำให้ได้

$$\left[\rho^2 C_y^2 + C_x^2 + 2\rho C_x C_y\right] > 0 \quad (4)$$

นั่นคือ ถ้าสมการข้างต้นเป็นจริงแล้ว ตัวประมาณ $\bar{y}_{pj}$ จะมีประสิทธิภาพสูงกว่าตัวประมาณ $\bar{y}_p$

3.2 เปรียบเทียบประสิทธิภาพของตัวประมาณ $\bar{y}_{pj}$ กับ $\bar{y}_{pr}$ เจริญฤทธิ์

เรามีเงื่อนไข

$$MSE_{\min}(\bar{y}_{pj}) - MSE(\bar{y}_p) < 0$$

$$\frac{1-f}{n} \bar{Y}^2 [1-\rho^2] C_y^2 - \frac{1-f}{n} \bar{Y}^2 [C_y^2 + \theta_1^2 C_x^2 + 2\theta_1 \rho C_x C_y] < 0$$

$$\frac{1-f}{n} \bar{Y}^2 [(1-\rho^2) C_y^2 - (C_y^2 + \theta_1^2 C_x^2 + 2\theta_1 \rho C_x C_y)] < 0$$

ทำให้ได้

$$-\left[\rho^2 C_y^2 + \theta_1^2 C_x^2 + 2\theta_1 \rho C_x C_y\right] < 0$$

ทำให้ได้

$$\left[\rho^2 C_y^2 + \theta_1 C_x (\theta_1 C_x + 2\rho C_y)\right] > 0 \quad (5)$$

นั่นคือ ถ้าสมการข้างต้นเป็นจริงแล้ว ตัวประมาณ จะมีประสิทธิภาพสูงกว่าตัวประมาณ

4. ตัวอย่างการคำนวณ

ผู้วิจัยใช้ข้อมูลของ Pandey และ Dubey (1988 อ้างถึงใน Khoshnevisan et al., [4]) ลักษณะของประชากร ประกอบไปด้วย

$$N = 20, n = 8, \bar{Y} = 19.55, \bar{X} = 18.8, C_x^2 = 0.1555, C_y^2 = 0.1262, \rho_{yx} = -0.9199$$

$$\text{และ } \beta_2 = 3.0613$$

ตรวจสอบเงื่อนไขข้างต้น

$$[\rho^2 C_y^2 + C_x^2 + 2\rho C_x C_y] = 0.004 > 0$$

และ

$$[\rho^2 C_y^2 + \theta_1 C_x (\theta_1 C_x + 2\rho C_y)] = 0.001 > 0$$

ซึ่งสอดคล้องกับเงื่อนไข

เกณฑ์การเปรียบเทียบ	ตัวประมาณ		
	แบบสามัญ ($\bar{y}_p$)	Singh et al. (2004) ($\bar{y}_{pr}$)	ที่เสนอ ($\bar{y}_{pi}$)
<i>MSE</i>	0.689	0.562	0.556
<i>PRE</i>	100	122.60	123.92
	81.57	100	101.08

จากตาราง จะเห็นว่าค่า MSE ของตัวประมาณที่นำเสนอมีค่าน้อยกว่า MSE ของตัวประมาณแบบผลคูณสามัญและของ Singh et al. [1] นั่นคือ ตัวประมาณที่นำเสนอมีประสิทธิภาพสูงกว่าตัวประมาณทั้งสอง เมื่อพิจารณาค่า PRE เทียบกับตัวประมาณ จะได้ค่า PRE ของตัวประมาณที่นำเสนอมีค่ามากกว่าตัวประมาณของ Singh et al. [1] เท่ากับ 1.32 ในขณะที่เทียบกับตัวประมาณของ [1] จะมีค่ามากกว่า 1.08

สรุปและอภิปรายผลงานวิจัย

ผู้วิจัยเสนอตัวประมาณแบบผลคูณตัวใหม่ของค่าเฉลี่ยประชากรในการสุ่มตัวอย่างอย่างง่ายพร้อมทั้งหาความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณที่นำเสนอและเปรียบเทียบประสิทธิภาพของตัวประมาณกับตัวประมาณแบบผลคูณสามัญและตัวประมาณ Singh และคณะ [1] ผลการวิจัยพบว่าตัวประมาณที่นำเสนอมีประสิทธิภาพสูงกว่าตัวประมาณแบบผลคูณสามัญและตัวประมาณของ Singh และคณะ [1] แบบมีเงื่อนไขการคำนวณเชิงตัวเลขให้ผลสอดคล้องกับผลเชิงทฤษฎี เมื่อเปรียบเทียบ PRE ของตัวประมาณที่นำเสนอเทียบกับตัวประมาณแบบผลคูณสามัญซึ่งมีค่าเท่ากับ 123.92 ในขณะที่ PRE ของ Singh และคณะ [1] เทียบกับตัวประมาณแบบผลคูณสามัญมีค่าเท่ากับ 122.60 จะเห็นว่ามีความสูงกว่าของ Singh และคณะ [1] ร้อยละ 1.32

เอกสารอ้างอิง

- [1] Singh, H.P., Tailor, R., and Kakaran, M.S. (2004). "An estimator of population mean using power transformation," **Journal of the Indian Society of Agricultural Statistics**. 58(2), 223-230.
- [2] Murthy, M.N. (1964). "Product method of estimation,". **Sankhya series A**. 26, 69-74.
- [3] Banerjee, J. and Tiwari, N. (2011). "Improved ratio type estimator using jack-knife method of Estimation," **Journal of reliability and statistical studies**. 4(1), 53-63.
- [4] Khoshnevisan, M., Singh, R., Chauhan, P., Sawan, N., and Smarandache, F. (2007). "A General family of estimators for estimating population mean using know value of some population parameter(s)," **Far East Journal of Theoretical Statistics**. 22(2), 181-191.
- [5] Singh, G.N. (2003). "On the improvement of product method of estimation in sample Surveys," **Journal of the Indian Society of Agricultural Statistics**. 56(3), 267-275.
- [6] Upadhyaya, L.N. and Singh, H.P. (1999). "Use of transformed auxiliary variable in estimating the finite population mean," **Biometrical Journal**. 41(5), 627-636.