

การประมาณค่าสัมประสิทธิ์การแปรผันเมื่อมีข้อมูลสูญหาย ที่เกิดจากการไม่ ตอบเฉพาะบางคำถาม

The Estimation of Coefficient of Variation in the Presence of Item Nonresponse

ชูเกียรติ โพนแก้ว*

สาขาวิชาคณิตศาสตร์และวิทยาการข้อมูล คณะวิทยาศาสตร์และเทคโนโลยี

มหาวิทยาลัยราชภัฏเพชรบูรณ์ เพชรบูรณ์ 67000

Chugiat Ponkaew*

Mathematics and Data Science Program, Faculty of Science and Technology,

Phetchabun Rajabhat University, Phetchabun 67000

บทคัดย่อ

การวิจัยในครั้งนี้มีวัตถุประสงค์เพื่อ นำเสนอตัวประมาณค่าสัมประสิทธิ์การแปรผัน กรณีที่มีข้อมูล สูญหาย ที่เกิดจากการไม่ตอบเฉพาะบางคำถาม ภายใต้การสุ่มแบบกำหนดความน่าจะเป็นให้เป็นสัดส่วนกับขนาดและใส่คืน รวมถึงสัดส่วนตัวอย่างมีขนาดเล็ก จากนั้นเปรียบเทียบตัวประมาณค่าสัมประสิทธิ์การแปรผันที่นำเสนอ กับตัวประมาณค่าสัมประสิทธิ์การแปรผันที่ปรับด้วยวิธีตัดข้อมูล ด้วยค่าความเอนเอียงสัมพัทธ์ และ ค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์

กำหนดความน่าจะเป็นที่จะตอบสนอง เท่ากับ 0.7 และ 0.85 กำหนดขนาดตัวอย่างเท่ากับ 15, 20, 25, 50, 120 และ 200 การวิจัยในครั้งนี้ใช้วิธีการจำลองแบบมอนติคาร์โล และกำหนดการทำซ้ำทั้งหมด 50,000 ครั้ง ในแต่ละระดับของความน่าจะเป็นที่จะตอบสนอง กรณีที่สัดส่วนตัวอย่างมีขนาดเล็ก ตัวประมาณค่าสัมประสิทธิ์การแปรผันที่นำเสนอ มีค่าความเอนเอียงสัมพัทธ์ และ ค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ ต่ำกว่า ตัวประมาณค่าสัมประสิทธิ์การแปรผันที่ปรับด้วยวิธีตัดข้อมูล

คำสำคัญ : ข้อมูลสูญหาย, ความน่าจะเป็นที่จะตอบสนอง, การจำลองแบบมอนติคาร์โล

* Corresponding author : chugiat.pon@pcru.ac.th

Received : 14/03/2022

Revised : 16/06/2022

Accepted : 07/09/2022

Abstract

The aim of this research is to proposed the estimator of coefficient of variation in the presence of item nonresponse and to compare the propose estimator with the adjusted coefficient of variation estimator by listwise deletion which was considered by the relative bias (RB) and relative root mean square error (RRMSE). Which were define the probability of response equal to 0.70 and 0.85 and set samples size were 15, 20, 25, 50, 120 and 200. And used the Mote Carlo simulation method with 50,000 times repeat per samples size.

The results of this research indicated that all probability of response and sampling fraction is negligible the and of the proposed estimator are lower than other estimator.

Keywords : Missing data, probability of response, Monte Carlo simulation

ที่มาและความสำคัญ

ค่าสัมประสิทธิ์การแปรผัน (Coefficient of Variation) เป็นหนึ่งในการวัดการกระจายสัมพัทธ์ ที่ใช้เปรียบเทียบการกระจายของข้อมูลตั้งแต่ 2 ชุดขึ้นไป เช่นใช้ในการเปรียบเทียบการกระจายของปริมาณข้าวโพดเลี้ยงสัตว์ที่ปลูกในเขตอำเภอหล่มเก่า และอำเภอหล่มสัก จังหวัดเพชรบูรณ์ เป็นต้น ค่าสัมประสิทธิ์การแปรผันมีค่าเท่ากับอัตราส่วนระหว่างส่วนเบี่ยงเบนมาตรฐาน และค่าเฉลี่ยประชากร ซึ่งโดยปกติจะไม่ทราบค่า เนื่องจากสูตรการคำนวณอยู่ในรูปฟังก์ชันของข้อมูลของประชากร (Population) ทำให้ต้องประมาณค่าจากตัวสถิติ (Statistic) ที่เป็นฟังก์ชันของข้อมูลของตัวอย่าง อย่างไรก็ตามกรณีที่มีข้อมูลสูญหาย (Missing data) ตัวสถิติดังกล่าวจะไม่สามารถใช้ประมาณค่าสัมประสิทธิ์การแปรผันได้เนื่องจากบางหน่วยของตัวอย่างไม่สามารถสังเกตค่าได้ ทำให้ผู้ศึกษาสนใจที่จะพัฒนา ตัวประมาณค่าของค่าสัมประสิทธิ์การแปรผัน ในกรณีที่มีข้อมูลสูญหาย ที่เกิดจากการไม่ตอบเฉพาะบางคำถาม (Item nonresponse)

กำหนดให้ y แทนตัวแปรที่สนใจศึกษา (Study variable) และ $U = \{1, 2, \dots, N\}$ แทนเซตของประชากรขนาด N กำหนดให้ k แทนตัวแปรช่วย (Auxiliary variable) ที่มีความสัมพันธ์กับตัวแปรที่สนใจศึกษา [1] ได้กำหนดสัญลักษณ์แทนค่าสัมประสิทธิ์การแปรผัน เขียนแทนด้วยสัญลักษณ์ θ_y และมีค่าตามที่แสดงในสมการ (1)

$$\theta_y = \sigma_y \bar{Y}^{-1} \quad (1)$$

เมื่อ $\sigma_y^2 = N^{-1} \sum_{i \in U} (y_i - \bar{Y})^2 = N^{-1} \sum_{i=1}^N (y_i - \bar{Y})^2$ แทนค่าความแปรปรวน และ $\bar{Y} = N^{-1} \sum_{i \in U} y_i = N^{-1} \sum_{i=1}^N y_i$ แทนค่าเฉลี่ยประชากร โดยปกติจะไม่สามารถคำนวณค่า κ เนื่องจากสูตรการคำนวณอ้างอิงถึงข้อมูลระดับประชากร ดังนั้นจำเป็นต้องประมาณค่า κ จากข้อมูลของตัวอย่าง สมมติว่าตัวอย่าง $s = \{j_1, j_2, \dots, j_n\}$ ขนาด

n เมื่อ $j_i \in U$ ถูกสุ่มตามข้อตกลงของแผนการสุ่มตัวอย่างแบบง่ายและใส่คืน (Simple random sampling with replacement: SRSWR) ตัวประมาณค่าของ θ_y เขียนแทนด้วย $\hat{\theta}_y$ มีค่าตามที่แสดงในสมการที่ (2)

$$\hat{\theta}_y = \hat{\sigma}_y \hat{Y}^{-1} \quad (2)$$

$$\text{เมื่อ } \hat{\sigma}_y = \sqrt{(n-1)^{-1} \sum_{i \in s} (y_i - \hat{Y})^2} = \sqrt{(n-1)^{-1} \sum_{i=1}^n (y_i - \hat{Y})^2} \text{ และ } \hat{Y} = n^{-1} \sum_{i \in s} y_i = \sum_{i=1}^n y_i$$

นอกจากนี้ [2-4] ได้นำเสนอตัวประมาณค่าสัมประสิทธิ์การแปรผัน เมื่อทราบข้อมูลของตัวแปรช่วยที่มีความสัมพันธ์กับตัวแปรที่สนใจศึกษา และ [5] ได้นำเสนอตัวประมาณค่าสัมประสิทธิ์การแปรผันกรณีที่ข้อมูลมีความสัมพันธ์กันในรูปอัตราส่วน

การสุ่มแบบ SRSWR จะพิจารณาแต่ละหน่วยของประชากรให้มีขนาดเท่ากัน ทำให้โอกาสที่แต่ละหน่วยของประชากร ถูกสุ่มมาเป็นตัวอย่างด้วยความน่าจะเป็นเท่ากัน แต่ในทางปฏิบัติขนาดของแต่ละหน่วยของประชากรจะมีขนาดไม่เท่ากัน เช่นตัวแปรที่สนใจศึกษาได้แก่ปริมาณของข้าวโพดเลี้ยงสัตว์ของเกษตรกรในอำเภอหล่มสักจังหวัดเพชรบูรณ์ โดยปกติปริมาณข้าวโพดเลี้ยงสัตว์ของเกษตรกรแต่ละคนจะมีค่าขึ้นอยู่กับพื้นที่ที่ใช้ในการเพาะปลูก ดังนั้นการกำหนดให้เกษตรกรแต่ละคนมีโอกาสที่จะถูกสุ่มด้วยความน่าจะเป็นเท่ากัน อาจทำให้การประมาณค่า σ_y และ $\bar{Y}$ เกิดความคลาดเคลื่อน ซึ่งจะส่งผลโดยตรงต่อการประมาณค่า θ_y กรณีที่ทราบข้อมูลของตัวแปรช่วย k ที่มีความสัมพันธ์กับตัวแปรที่สนใจศึกษา เช่นทราบพื้นที่ที่เกษตรกรใช้เพาะปลูกข้าวโพดเลี้ยงสัตว์ การประมาณค่า σ_y และ $\bar{Y}$ นิยมศึกษาภายใต้การสุ่มแบบกำหนดความน่าจะเป็นให้เป็นสัดส่วนกับขนาด และใส่คืน (Probability Proportional to Size and with Replacement: PPSWR) ซึ่งภายใต้การสุ่มดังกล่าวค่าความน่าจะเป็นที่หน่วยตัวอย่างที่ i เมื่อ $i \in U$ จะถูกสุ่มมาเป็นตัวอย่างมีค่าขึ้นอยู่กับค่า k_i กล่าวคือถ้า k_i มีค่ามาก โอกาสที่หน่วยตัวอย่างที่ i จะถูกสุ่มมาเป็นตัวอย่างก็จะมีค่ามาก แต่ถ้าค่า k_i มีค่าน้อย โอกาสที่หน่วยตัวอย่างที่ i จะถูกสุ่มมาเป็นตัวอย่างก็จะมีค่าน้อย ภายใต้การสุ่มแบบ PPSWR [6] นำเสนอตัวประมาณค่าเฉลี่ยประชากร ตามที่แสดงในสมการที่ (3)

$$\hat{Y}_{HH} = N^{-1} \hat{Y}_{HH} = (nN)^{-1} \sum_{i \in s} y_i P_i^{-1} = (nN)^{-1} \sum_{i=1}^n y_i P_i^{-1} \quad (3)$$

เมื่อ $\hat{Y}_{HH} = n^{-1} \sum_{i \in s} y_i P_i^{-1} = n^{-1} \sum_{i=1}^n y_i P_i^{-1}$ และ $P_i = k_i K^{-1}$ แทนความน่าจะเป็นที่หน่วยตัวอย่างที่ i จะถูกสุ่มมาเป็นตัวอย่าง โดยที่ $K = \sum_{i \in U} k_i$ สำหรับค่าความแปรปรวนของ $\hat{Y}_{HH}$ เขียนแทนด้วย $V(\hat{Y}_{HH})$ และมีค่าตามที่แสดงในสมการที่ (4)

$$V(\hat{Y}_{HH}) = N^{-2} V(\hat{Y}_{HH}) = (N^2 n)^{-1} \sum_{i \in U} P_i (y_i P_i^{-1} - Y)^2 \quad (4)$$

$$\text{เมื่อ } V(\hat{Y}_{HH}) = n^{-1} \sum_{i \in U} P_i (y_i P_i^{-1} - Y)^2 = n^{-1} \sum_{i=1}^N P_i (y_i P_i^{-1} - Y)^2 \text{ และ } Y = \sum_{i \in U} y_i = \sum_{i=1}^N y_i$$

นอกจากนี้ตัวประมาณค่าของ $V(\hat{Y}_{HH})$ เขียนแทนด้วย $\hat{V}(\hat{Y}_{HH})$ และมีค่าตามที่แสดงในสมการที่ (5)

$$\hat{V}(\hat{Y}_{HH}) = (N^{-2}n(n-1))^{-1} \sum_{i \in s} (y_i P_i^{-1} - \hat{Y}_{HH})^2 = (N^{-2}n(n-1))^{-1} \sum_{i=1}^n (y_i P_i^{-1} - \hat{Y}_{HH})^2 \quad (5)$$

ต่อมากรณีที่ไม่มีการสุ่มตัวอย่าง [7] ได้นำเสนอตัวประมาณค่าของ κ ภายใต้การสุ่มตัวอย่างแบบ PPSWR เขียนแทนด้วยสัญลักษณ์ $\hat{\kappa}_{HH}$ และมีค่าตามที่แสดงในสมการที่ (6)

$$\hat{\kappa}_{HH} = \hat{\sigma}_{yHH} \hat{Y}_{HH}^{-1} \quad (6)$$

$$\begin{aligned} \text{เมื่อ } \hat{\sigma}_{yHH} &= \left[(Nn)^{-1} \sum_{i \in s} y_i^2 P_i^{-1} - \hat{Y}_{HH}^2 + \hat{V}(\hat{Y}_{HH}) \right]^{\frac{1}{2}} \\ &= \left[(Nn)^{-1} \sum_{i=1}^n y_i^2 P_i^{-1} - \hat{Y}_{HH}^2 + \hat{V}(\hat{Y}_{HH}) \right]^{\frac{1}{2}} \end{aligned}$$

อย่างไรก็ตามกรณีที่ตัวแปรที่สนใจศึกษามีข้อมูลสูญหาย (Nonresponse) ที่เกิดจากการไม่ตอบเฉพาะบางคำถาม ตัวประมาณค่าที่ [7] นำเสนอไม่สามารถใช้ประมาณค่า κ ได้เนื่องจากบางหน่วยตัวอย่างในตัวอย่าง s ขนาด n ไม่สามารถสังเกตค่าได้ ซึ่งในกรณีเช่นนี้อาจแก้ปัญหาก็ได้โดยใช้วิธีการตัดข้อมูล (ตัดหน่วยตัวอย่างที่ไม่สามารถสังเกตค่าได้) และคำนวณค่าเฉพาะเซตตัวอย่างที่สามารถสังเกตค่าได้ แต่วิธีการดังกล่าวอาจทำให้ประสิทธิภาพของตัวประมาณค่าลดลงเมื่อข้อมูลสูญหายมีขนาดใหญ่กรณีที่ไม่มีข้อมูลสูญหาย ต่อมา [8] นำเสนอตัวประมาณค่าเฉลี่ยประชากร ภายใต้การสุ่มเดียวกับ [7] แต่พิจารณากรณีที่ไม่มีข้อมูลสูญหายที่เกิดจากการไม่ตอบเฉพาะบางคำถาม และมีค่าตามที่แสดงในสมการที่ (7)

$$\hat{Y}_{IPPS} = \left(\sum_{i \in s} \frac{1}{\pi_i} \sum_{i \in s} \frac{r_i y_i}{\pi_i} \right) \left(N \sum_{i \in s} \frac{r_i}{\pi_i} \right)^{-1} = \left(\sum_{i=1}^n \frac{1}{\pi_i} \sum_{i=1}^n \frac{r_i y_i}{\pi_i} \right) \left(N \sum_{i=1}^n \frac{r_i}{\pi_i} \right)^{-1} \quad (7)$$

เมื่อ $\pi_i = p(i \in s) = nP_i$ ค่า $r_i = 1$ เมื่อหน่วยตัวอย่างที่ i สามารถสังเกตค่าได้ และ $r_i = 0$ เมื่อหน่วยตัวอย่างที่ i สูญหาย นอกจากนี้ภายใต้เฟรมเวิร์คแบบย้อนกลับ (Reverse framework) และสัดส่วนตัวอย่างมีขนาดเล็ก [8] ได้นำเสนอ ตัวประมาณค่าความแปรปรวนของตัวประมาณค่าเฉลี่ยประชากร ตามที่แสดงในสมการที่ (8)

$$\begin{aligned} \hat{V}(\hat{Y}_{IPPS}) &= \left(N \sum_{i \in s} \frac{1}{\pi_i^{-1}} \right)^{-2} n(n-1)^{-1} \left[\sum_{i \in s} \hat{z}_i^2 - n^{-1} \left(\sum_{i \in s} \hat{z}_i \right)^2 \right] \\ &= \left(N \sum_{i=1}^n \frac{1}{\pi_i^{-1}} \right)^{-2} n(n-1)^{-1} \left[\sum_{i=1}^n \hat{z}_i^2 - n^{-1} \left(\sum_{i=1}^n \hat{z}_i \right)^2 \right] \end{aligned} \quad (8)$$

$$\begin{aligned}\text{เมื่อ } \hat{z}_i &= \sum_{i \in s} r_i y_i \pi_i^{-1} + \left(N \sum_{i \in s} \pi_i^{-1} \right) r_i \left[y_i - \sum_{i \in s} r_i y_i \pi_i^{-1} \left(\sum_{i \in s} r_i \pi_i^{-1} \right)^{-1} \right] \\ &= \sum_{i=1}^n r_i y_i \pi_i^{-1} + \left(N \sum_{i=1}^n \pi_i^{-1} \right) r_i \left[y_i - \sum_{i=1}^n r_i y_i \pi_i^{-1} \left(\sum_{i=1}^n r_i \pi_i^{-1} \right)^{-1} \right]\end{aligned}$$

ดังนั้นการวิจัยในครั้งนี้มีวัตถุประสงค์เพื่อ (1) นำเสนอตัวประมาณค่าสัมประสิทธิ์การแปรผัน เมื่อมีข้อมูลสูญหาย โดยใช้ตัวประมาณค่าของ [8] ปรับตัวประมาณค่าของ [7] และ (2) เปรียบเทียบตัวประมาณค่าสัมประสิทธิ์การแปรผันที่นำเสนอ กับตัวประมาณค่าสัมประสิทธิ์การแปรผันที่ปรับด้วยวิธีตัดข้อมูล ด้วยค่าความเอนเอียงสัมพัทธ์ และ ค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ ทั้งนี้ผู้วิจัยกำหนดข้อตกลงเบื้องต้นที่ใช้ในการประมาณค่าสัมประสิทธิ์การแปรผัน ดังนี้ (A₁) การสูญหายของข้อมูลเกิดขึ้นเฉพาะตัวแปรที่สนใจศึกษา และเป็นแบบไม่ตอบเฉพาะบางคำถาม และ (A₂) สัดส่วนตัวอย่างมีขนาดเล็ก

วิธีการดำเนินการวิจัย

1. แผนการสุ่มตัวอย่าง

การวิจัยในครั้งนี้ศึกษาเกี่ยวกับการประมาณค่าสัมประสิทธิ์การแปรผันภายใต้แผนการสุ่มตัวอย่างแบบ กำหนดความน่าจะเป็นให้เป็นสัดส่วนกับขนาด และใส่คืน หรือแบบ PPSWR ซึ่งแผนการสุ่มตัวอย่างดังกล่าวจะกำหนดความน่าจะเป็นที่แต่ละหน่วยตัวอย่างของประชากรจะถูกสุ่มมาเป็นตัวอย่าง ตามขนาดของตัวแปรช่วยที่มีความสัมพันธ์กับตัวแปรที่สนใจศึกษา [9] ได้อธิบายวิธีการสุ่มตัวอย่างของแผนการสุ่มแบบ PPSWR ดังนี้

กำหนดให้ $B_0 = \max\{k_1, k_2, \dots, k_N\}$ และ $B \geq B_0$

ขั้นที่ 1 สุ่มจำนวนเต็ม i เมื่อ $1 \leq i \leq N$

ขั้นที่ 2 สุ่มจำนวนเต็ม R เมื่อ $1 \leq R \leq B$

ขั้นที่ 3 ถ้า $R \leq k_i$ แล้ว หน่วยตัวอย่างที่ i ถูกเลือกเป็นสมาชิกของตัวอย่าง s

ขั้นที่ 4 ทำซ้ำขั้นที่ 1 -3 จนกระทั่งได้ตัวอย่าง s ขนาด n

2. นำเสนอตัวประมาณค่าสัมประสิทธิ์การแปรผัน

การวิจัยในครั้งนี้มีวัตถุประสงค์ เพื่อนำเสนอตัวประมาณค่าสัมประสิทธิ์การแปรผัน กรณีที่มีข้อมูลสูญหายที่เกิดจากการไม่ตอบเฉพาะบางคำถาม ขั้นแรกผู้วิจัยจะได้กำหนดนิยามศัพท์เพื่อใช้ในการอ้างอิงตัวประมาณค่าสัมประสิทธิ์การแปรผันดังนี้

ตัวประมาณค่าสัมประสิทธิ์การแปรผันแบบตัดข้อมูล หมายถึง ตัวประมาณค่าที่ได้จากการปรับตัวประมาณค่าของ [7] ด้วยวิธีตัดข้อมูล

ตัวประมาณค่าสัมประสิทธิ์การแปรผันที่นำเสนอ หมายถึง ตัวประมาณค่าที่ได้จากการปรับตัวประมาณค่าของ [7] ด้วยตัวประมาณค่าของ [8]

ต่อไปผู้วิจัยจะได้กล่าวถึงขั้นตอนของการนำเสนอตัวประมาณค่าดังนี้

การประมาณค่าสัมประสิทธิ์การแปรผัน ภายใต้การสุ่มแบบ PPSWR และไม่มีการสูญหายของข้อมูล [7] นำเสนอตัวสถิติที่อยู่ในรูปฟังก์ชันของ y_i และ k_i เมื่อ $i \in s$ เพื่อใช้ประมาณค่าสัมประสิทธิ์การแปรผัน และมีค่าตามที่แสดงในสมการที่ (6) กรณีที่มีการสูญหายของข้อมูล ตัวอย่าง s ขนาด n ถูกแบ่งออกเป็นเซตที่สามารถสังเกตค่า y_i ได้ เขียนแทนด้วยเซต s_r และเซต s_m ซึ่งไม่สามารถสังเกตค่า y_i ได้ โดยที่ $s_r \cup s_m = s$ และ $s_r \cap s_m = \emptyset$ นอกจากนี้ เซต s_r และเซต s_m มีขนาดเท่ากับ n_r และ n_m ตามลำดับ ดังนั้นกรณีที่เกิดการสูญหายของข้อมูล ตัวประมาณค่าของ [7] ไม่สามารถใช้ประมาณค่าสัมประสิทธิ์การแปรผันได้เนื่องจากค่า y_i เมื่อ $i \in s_m$ ไม่สามารถสังเกตค่าได้ วิธีการแก้ปัญหาเมื่อเกิดการสูญหายของข้อมูล นิยมใช้วิธีตัดข้อมูลแบบ Listwise วิธีการดังกล่าวจะตัดค่า y_i ที่ไม่สังเกตค่าได้ออกจากตัวอย่าง s ขนาด n กล่าวคือใช้เฉพาะค่า y_i เมื่อ $i \in s_r$ ในการประมาณค่าพารามิเตอร์ที่ต้องการศึกษา โดยอาศัยแนวคิดดังกล่าว การประมาณค่าสัมประสิทธิ์การแปรผัน กรณีที่มีการสูญหายของข้อมูล ด้วยตัวประมาณค่าของ [7] มีค่าตามที่แสดงในสมการที่ (9)

$$\hat{\kappa}_{HHsr} = \hat{\sigma}_{yHHsr} \hat{Y}_{yHHsr}^{-1} \quad (9)$$

$$\begin{aligned} \text{เมื่อ} \quad \hat{\sigma}_{yHHsr} &= \left[(Nn)^{-1} \sum_{i \in s_r} y_i^2 P_i^{-1} - \hat{Y}_{yHHsr}^2 + \hat{V}(\hat{Y}_{yHHsr}) \right]^{-\frac{1}{2}} \\ &= \left[(Nn)^{-1} \sum_{i=1}^{n_r} y_i^2 P_i^{-1} - \hat{Y}_{yHHsr}^2 + \hat{V}(\hat{Y}_{yHHsr}) \right]^{-\frac{1}{2}} \\ \text{ค่า} \quad \hat{Y}_{yHHsr} &= (n_r N)^{-1} \sum_{i=1}^{n_r} y_i P_i^{-1} \\ \text{และ} \quad \hat{V}(\hat{Y}_{yHHsr}) &= (N^{-2} n_r (n_r - 1))^{-1} \sum_{i \in s_r} (y_i P_i^{-1} - \hat{Y}_{HHsr})^2 \\ &= (N^{-2} n_r (n_r - 1))^{-1} \sum_{i=1}^{n_r} (y_i P_i^{-1} - \hat{Y}_{HHsr})^2 \end{aligned}$$

การประมาณค่าโดยใช้วิธีตัดข้อมูลแบบ Listwise เป็นวิธีที่สะดวก แต่อาจส่งผลกระทบต่อประสิทธิภาพของตัวประมาณค่าในกรณีที่ข้อมูลสูญหายมีขนาดใหญ่ ดังนั้นการวิจัยในครั้งนี้ผู้วิจัยต้องการพัฒนาตัวประมาณค่าของ [7] โดยใช้ตัวประมาณค่าของ [8] ปรับตัวประมาณค่าดังกล่าว และเขียนแทนตัวประมาณค่าที่นำเสนอด้วยสัญลักษณ์ $\hat{\kappa}_{CP}$

3. แสดงตัวอย่างเชิงตัวเลข

3.1 สร้างข้อมูลของประชากรขนาด $N = 500$ โดยใช้ตัวแบบตามที่แสดงในสมการที่ (10)

$$y_i = \beta_0 + \beta_1 k_i + \varepsilon_i \quad (10)$$

เมื่อ $(\beta_0, \beta_1) = (9100, 250)$, $k_i \sim N(80, 25)$ และ $\varepsilon_i \sim N(0, 5)$

3.2 กำหนดการสูญหายของข้อมูลหรือกำหนดค่า r_i เมื่อ $i \in U$

(i) กำหนดค่าความน่าจะเป็นที่จะตอบสนองทั้งหมด 2 ระดับ ประกอบด้วย $p = 0.7$ และ $p = 0.85$

(ii) แต่ระดับของค่า p เปรียบเทียบค่า p กับค่า A_i เมื่อ $A_i \sim U(0, 1)$ ถ้าค่า $A_i \leq p$ แล้วค่า $r_i = 1$ กรณีอื่น ๆ ค่า $r_i = 0$

3.3 แต่ละระดับของค่า p สุ่มตัวอย่าง s ขนาด n ทั้งหมด 6 ระดับ ประกอบด้วย $n = 15, 20, 25, 50, 120$ และ 200 โดยใช้แผนการสุ่มตัวอย่างแบบ PPSWR และทำซ้ำทั้งหมด 50,000 รอบ ($M = 50,000$)

(i) ที่ $p = 0.7$ และ $n = 15$ กำหนด $m = 1$ แล้วคำนวณค่า $\hat{\kappa}_{HHsr[m]}$ และ $\hat{\kappa}_{CP[m]}$

(ii) กำหนดให้ $m = m + 1$ แล้วทำซ้ำขั้นที่ (i) และ (ii) จนกระทั่ง $m = 50,000$

(iii) คำนวณค่าความเอนเอียงสัมพัทธ์ (Relative bias) และค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ (Relative root mean square error) ของตัวประมาณค่าสัมประสิทธิ์การแปรผัน ตามที่แสดงในสมการที่ (11) และสมการที่ (12) ดังนี้

$$PB(\hat{\kappa}_{(i)}) = \frac{|E(\hat{\kappa}_{(i)}) - \kappa|}{\kappa} \quad ; i = 1, 2 \quad (11)$$

$$PRMSE(\hat{\kappa}_{(i)}) = \frac{\sqrt{\frac{1}{M-1} \sum_{m=1}^M (\hat{\kappa}_{(i[m])} - \kappa)^2}}{\kappa} \quad ; i = 1, 2 \quad (12)$$

เมื่อ $\hat{\kappa}_{(i[m])} = \hat{\kappa}_{(1)}$ หรือ $\hat{\kappa}_{(2)}$ ค่า $\hat{\kappa}_{(1)} = \hat{\kappa}_{HHsr}$ ค่า $\hat{\kappa}_{(2)} = \hat{\kappa}_{CP}$ นอกจากนี้ค่า $\hat{\kappa}_{(i[m])}$ หมายถึงค่าของ $\hat{\kappa}_{(i)}$ ในรอบที่ m

(iv) ทำซ้ำขั้นตอนที่ (i)-(iii) แต่เปลี่ยนค่า n จาก 15 เป็น 20, 25, 50, 120 และ 200

(v) ทำซ้ำขั้นตอนที่ (i)-(v) แต่เปลี่ยนค่า p จาก 0.70 เป็น 0.85

3.4 เปรียบเทียบประสิทธิภาพของตัวประมาณค่าสัมประสิทธิ์การแปรผัน

บทนิยาม 1 กำหนดให้ $\hat{\kappa}_A$ และ $\hat{\kappa}_B$ แทนตัวประมาณค่าของสัมประสิทธิ์การแปรผัน ค่า $RRMSE(\hat{\kappa}_A)$ และ $RRMSE(\hat{\kappa}_B)$ แทนค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ ของ $\hat{\kappa}_A$ และ $\hat{\kappa}_B$ ค่า $RB(\hat{\kappa}_A)$ และ $RB(\hat{\kappa}_B)$ แทนค่าความเอนเอียงสัมพัทธ์ ของ $\hat{\kappa}_A$ และ $\hat{\kappa}_B$ ตามลำดับ ถ้า $RRMSE(\hat{\kappa}_A) < RRMSE(\hat{\kappa}_B)$ และ $RB(\hat{\kappa}_A) < RB(\hat{\kappa}_B)$ แล้ว $\hat{\kappa}_A$ มีประสิทธิภาพดีกว่า $\hat{\kappa}_B$

ผลการวิจัยและอภิปรายผลการวิจัย

1. ผลการวิจัย

การวิจัยในครั้งนี้มีวัตถุประสงค์เพื่อนำเสนอตัวประมาณค่าสัมประสิทธิ์การแปรผัน ภายใต้แผนการสุ่มตัวอย่างแบบ PPSWR และตัวแปรที่สนใจศึกษาเกิดการสูญหายของข้อมูลแบบไม่ตอบเฉพาะบางคำถามรวมถึงสัดส่วนตัวอย่างมีขนาดเล็ก ค่าสัมประสิทธิ์การแปรผัน เป็นหนึ่งในการวัดการกระจายสัมพัทธ์และใช้เปรียบเทียบการกระจายของข้อมูลมีค่าเท่ากับอัตราส่วนระหว่างส่วนเบี่ยงเบนมาตรฐานกับค่าเฉลี่ยเลขคณิต การเปรียบเทียบการกระจายของข้อมูลที่เป็นประชากร สูตรการคำนวณค่าสัมประสิทธิ์การแปรผัน จะเขียนอยู่ในรูปฟังก์ชันค่าสังเกตของตัวแปรที่สนใจศึกษาทุกค่าที่เป็นหน่วยตัวอย่างของประชากร กล่าวคือจะใช้ค่า y_i ทุกค่า $i \in U$ กรณีที่ไม่สามารถสังเกตค่า y_i ได้ทุกค่า $i \in U$ จำเป็นต้องสุ่มตัวอย่างและใช้ตัวสถิติที่คำนวณได้จากตัวอย่าง s ขนาด n ในการประมาณค่าสัมประสิทธิ์การแปรผัน ภายใต้การสุ่มแบบ PPSWR และทราบข้อมูลของตัวแปรช่วยที่มีความสัมพันธ์กับตัวแปรที่สนใจศึกษา รวมถึงไม่มีการสูญหายของข้อมูล [6] ได้นำเสนอตัวประมาณค่าเฉลี่ยประชากร และตัวประมาณค่าความแปรปรวนของตัวประมาณค่าดังกล่าว ต่อมา [7] ได้ปรับตัวประมาณค่าของ [6] เพื่อใช้ประมาณค่าสัมประสิทธิ์การแปรผัน ตามที่แสดงในสมการที่ (6) ซึ่งเขียนอยู่ในรูปฟังก์ชันของตัวสถิติ 3 ค่า ประกอบด้วย ค่า $\hat{Y}_{HH}$ ค่า $\hat{V}(\hat{Y}_{HH})$ และค่า $\hat{T} = (Nn)^{-1} \sum_{i \in s} y_i^2 P_i^{-1} = (Nn)^{-1} \sum_{i=1}^{n_r} y_i^2 P_i^{-1}$ เนื่องจากค่าสถิติทั้ง 3 ค่าจำเป็นต้องใช้ค่าสังเกตของตัวแปรที่สนใจศึกษา หรือ y_i ทุกค่า $i \in s$ ส่งผลให้กรณีที่บางหน่วยตัวอย่างที่เป็นสมาชิกของตัวอย่าง s ขนาด n ไม่สามารถสังเกตค่า y_i ได้ ตัวประมาณค่าที่ [7] นำเสนอจึงไม่สามารถประมาณค่าสัมประสิทธิ์การแปรผันได้

กรณีที่ข้อมูลสูญหายที่เกิดจากการไม่ตอบเฉพาะบางคำถาม [8] ได้นำเสนอตัวประมาณค่าเฉลี่ยประชากร และตัวประมาณค่าความแปรปรวนของตัวประมาณค่าดังกล่าว เขียนแทนด้วยสัญลักษณ์ $\hat{Y}_{IPPS}$ และ $\hat{V}(\hat{Y}_{IPPS})$ ตามลำดับ ดังนั้นผู้วิจัยได้ทำการพัฒนาตัวประมาณค่าของ [1] โดยการแทนค่า ตัวสถิติ $\hat{Y}_{HH}$ ด้วย $\hat{Y}_{IPPS}$ และแทนค่าตัวสถิติ $\hat{V}(\hat{Y}_{HH})$ ด้วย $\hat{V}(\hat{Y}_{IPPS})$ สำหรับค่า $\hat{T}$ ผู้วิจัยดำเนินการปรับค่าดังนี้ กำหนดค่า $y_s = \{y_1, y_2, \dots, y_n\}$ แทนเซตของค่าสังเกตของตัวอย่าง s ขนาด n เซต s_r แทนเซตของตัวอย่างที่สามารถสังเกตค่า y_i และ s_m แทนเซตที่ไม่สามารถสังเกตค่า y_i ได้ เมื่อ $s_r = \{y_1, y_2, \dots, y_{nr}\}$ และ $s_m = \{y_{nr+1}, y_{nr+2}, \dots, y_n\}$ โดยที่ $s_r \cup s_m = s$ และ $s_r \cap s_m = \emptyset$ ภายใต้ข้อตกลง (A_1) และ (A_2) [8] ใช้วิธีแทนค่าข้อมูลสูญหายด้วยค่าเฉลี่ย $y^* = \sum_{i \in s} d_i r_i y_i / \sum_{i \in s} d_i r_i = \sum_{i=1}^n d_i r_i y_i / \sum_{i=1}^n d_i r_i$ เมื่อ $d_i = \pi_i^{-1}$ กล่าวคือแทนค่า y_i เมื่อ $i \in s_m$ ด้วยค่า y^* หลังจากแทนค่าข้อมูลสูญหายด้วยค่าเฉลี่ย สามารถปรับเซตของ y_s เขียนแทนด้วยสัญลักษณ์ y_s^* และมีค่าตามที่แสดงในสมการที่ (13)

$$y_s^* = \{y_i^* \mid y_i^* = y_i \text{ เมื่อ } r_i = 1 \text{ หรือ } y_i^* = y^* \text{ เมื่อ } r_i = 0\} \quad (13)$$

ต่อไปใช้เซตของค่า y_s^* ตามที่แสดงในสมการ (13) ปรับค่า $\hat{T}$ เขียนแทนด้วย $\hat{T}_{CP}^*$ และมีค่าตามที่แสดงในสมการ (14)

$$\hat{T}_{CP}^* = (Nn)^{-1} \sum_{i \in s} \frac{1}{P_i} [r_i y_i^2 + (1 - r_i)(y_i^*)^2] = (Nn)^{-1} \sum_{i \in s} \frac{1}{P_i} (y_i^*)^2 = (Nn)^{-1} \sum_{i=1}^n \frac{1}{P_i} (y_i^*)^2 \quad (14)$$

ผู้วิจัยทำการปรับค่า $\hat{\sigma}_{yHHsr}$ เขียนแทนด้วย $\hat{\sigma}_{CP}$ และกำหนดค่าตามที่แสดงในสมการที่ (15)

$$\hat{\sigma}_{CP} = \left[\hat{T}_{CP}^* (Nn)^{-1} - \hat{Y}_{IPPS}^2 + \hat{V}(\hat{Y}_{IPPS}) \right]^{\frac{1}{2}} \quad (15)$$

สมการที่ (14) และ (15) ผู้วิจัยได้นำเสนอค่า $\hat{T}_{CP}^*$ และค่า $\hat{\sigma}_{CP}$ สำหรับตัวประมาณค่าสัมประสิทธิ์การแปรผันที่ผู้วิจัยนำเสนอแสดงดังบทตั้งที่ 2

บทตั้ง 2 กำหนดให้ข้อตกลง (A_1) และ (A_2) เป็นจริง ภายใต้การสุ่มแบบ PPSWR ตัวประมาณค่าสัมประสิทธิ์การแปรผันที่ผู้วิจัยนำเสนอ กำหนดค่าดังนี้ $\hat{k}_{CP} = \hat{\sigma}_{CP} \hat{Y}_{CP}^{-1}$ เมื่อค่า $\hat{\sigma}_{CP}$ กำหนดตามสมการ (15) และค่า $\hat{Y}_{CP} = \hat{Y}_{IPPS}$

การประมาณค่าสัมประสิทธิ์การแปรผัน กรณีที่มีข้อมูลสูญหายที่เกิดจากการไม่ตอบเฉพาะบางคำถาม และ สัดส่วนตัวอย่างมีขนาดเล็ก ภายใต้การสุ่มแบบ PPSWR ผู้วิจัยได้นำเสนอตัวประมาณค่าตามที่แสดงในบทตั้งที่ 2 ขึ้นต่อไปผู้วิจัยจะได้นำเสนอค่าที่ใช้เปรียบเทียบประสิทธิภาพของตัวประมาณค่าสัมประสิทธิ์การแปรผันที่นำเสนอ กับตัวประมาณค่าสัมประสิทธิ์การแปรผันแบบตัดข้อมูล ประกอบด้วย ค่าความเอนเอียงสัมพัทธ์ และค่ารากที่สองของความคลาดเคลื่อนกำลังสองสัมพัทธ์ ตามที่แสดงในตารางที่ 1 ผลที่ได้จากตารางพบว่า ในแต่ละระดับของค่าความน่าจะเป็นที่จะตอบสนอง และสัดส่วนตัวอย่างมีขนาดเล็ก ($n \leq 50$) ค่าที่ใช้เปรียบเทียบประสิทธิภาพของตัวประมาณค่าที่นำเสนอมีค่าต่ำกว่าค่าที่ได้จากตัวประมาณค่าสัมประสิทธิ์การแปรผันแบบตัดข้อมูล โดยอาศัยบทนิยามที่ 1 ได้ว่าตัวประมาณค่าสัมประสิทธิ์การแปรผันที่นำเสนอ มีประสิทธิภาพดีกว่าตัวประมาณค่าสัมประสิทธิ์การแปรผันแบบตัดข้อมูล กรณีที่ขนาดตัวอย่างไม่มีขนาดเล็ก ($n = 120$ และ $n = 200$) ผลการเปรียบเทียบประสิทธิภาพ ไม่สามารถสรุปได้แน่นอน เนื่องจากค่าความเอนเอียงสัมพัทธ์ และค่ารากที่สองของความคลาดเคลื่อนกำลังสองสัมพัทธ์ ไม่สอดคล้องกับบทนิยามที่ 1

2. อภิปรายผลการวิจัย

ผลการเปรียบเทียบประสิทธิภาพในตารางที่ 1 พบว่า กรณีที่ขนาดตัวอย่างเป็นไปตามข้อตกลงเบื้องต้น กล่าวคือ สัดส่วนตัวอย่างมีขนาดเล็ก ($n \leq 50$) ตัวประมาณค่าที่นำเสนอมีประสิทธิภาพดีกว่าตัวประมาณค่าแบบตัดข้อมูลสาเหตุเนื่องจากตัวประมาณค่าตัวประสิทธิ์การแปรผันที่คำนวณตามสูตรของ [7] ใช้วิธีการตัดข้อมูลแบบ Listwise

ตารางที่ 1: ค่ารากที่สองของความคลาดเคลื่อนกำลังสองสัมพัทธ์ และค่าความเอนเอียงสัมพัทธ์ ของ $\hat{\kappa}_{HHsr}$ และ $\hat{\kappa}_{CP}$

p	n	$RB(\hat{\kappa}_{HHsr})$	$RB(\hat{\kappa}_{CP})$	$RRMSE(\hat{\kappa}_{HHsr})$	$RRMSE(\hat{\kappa}_{CP})$
0.70	15	0.3186	0.2030	0.6428	0.5467
	20	0.2431	0.1284	0.5767	0.4903
	25	0.1940	0.0785	0.5347	0.4588
	50	0.0728	0.0453	0.4420	0.3935
	120	0.0363	0.0373	0.3624	0.3629
	200	0.0350	0.0348	0.3451	0.3163
0.85	15	0.2636	0.2109	0.5929	0.5496
	20	0.2038	0.1475	0.5413	0.5034
	25	0.1535	0.1007	0.5040	0.4681
	50	0.0473	0.0123	0.4191	0.3970
	120	0.0156	0.0116	0.3422	0.3448
	200	0.0148	0.0010	0.2994	0.3099

กล่าวคือตัดหน่วยตัวอย่างที่ไม่สามารถสังเกตค่าได้ออกไปจากตัวอย่าง s ขนาด n และใช้เฉพาะหน่วยตัวอย่างที่สามารถสังเกตค่าได้ในเซต s_i ขนาด n_i ในการคำนวณตัวประมาณค่าสัมประสิทธิ์การแปรผัน อย่างไรก็ตามกรณีที่ข้อมูลสูญหายมีขนาดใหญ่ วิธีการดังกล่าวอาจส่งผลให้ประสิทธิภาพของตัวประมาณค่า รวมถึงอำนาจการทดสอบเชิงสถิติลดลง นอกจากนี้ [10] กล่าวว่า การประมาณค่าในกรณีที่ข้อมูลสูญหาย บางครั้งไม่สามารถใช้วิธีตัดข้อมูลแบบ Listwise เนื่องจากข้อมูลที่สนใจศึกษามีจำนวนไม่มาก ทำให้ต้องมีการประมาณค่าข้อมูลสูญหาย ต่อไปพิจารณากรณีที่ขนาดตัวอย่างไม่สอดคล้องกับข้อตกลงเบื้องต้น ผลการเปรียบเทียบประสิทธิภาพ ไม่สามารถสรุปได้แน่นอน เนื่องจากค่าความเอนเอียงสัมพัทธ์ และค่ารากที่สองของความคลาดเคลื่อนกำลังสองสัมพัทธ์ ไม่สอดคล้องกับบทนิยามที่ 1 สาเหตุอาจเป็นผลมาจากตัวประมาณค่าสัมประสิทธิ์การแปรผันที่ผู้วิจัยนำเสนอ ใช้ตัวประมาณค่าของ [8] ปรับตัวประมาณค่าของ [7] แทนวิธีตัดข้อมูลแบบ Listwise ทั้งสองท่านได้นำเสนอตัวประมาณค่าเฉลี่ยประชากร และตัวประมาณค่าความแปรปรวน โดยใช้วิธีแทนที่ข้อมูลสูญหายด้วยค่าเฉลี่ย ผลการเปรียบเทียบประสิทธิภาพพบว่า ตัวประมาณค่าที่ได้จากการแทนที่ข้อมูลสูญหายด้วยค่าเฉลี่ยมีประสิทธิภาพดีกว่าตัวประมาณค่าที่ได้จากการใช้วิธีตัดข้อมูลแบบ Listwise เฉพาะกรณีที่สัดส่วนตัวอย่างมีขนาดเล็ก ($n \leq 50$)

สรุปผลการวิจัย

การประมาณค่าสัมประสิทธิ์การแปรผัน กรณีที่ตัวแปรที่สนใจศึกษาเกิดการสูญหายของข้อมูลแบบไม่ตอบเฉพาะบางคำถาม ภายใต้แผนการสุ่มตัวอย่างแบบ PPSWR และพิจารณาเฉพาะกรณีที่สัดส่วนตัวอย่างมีขนาดเล็ก ตัวประมาณค่าสัมประสิทธิ์การแปรผันที่ผู้วิจัยนำเสนอคือ

$$\hat{k}_{CP} = \hat{\sigma}_{CP} \hat{Y}_{CP}^{-1} \quad (16)$$

$$\text{เมื่อ } \hat{\sigma}_{CP} = \left[(Nn)^{-1} \hat{T}_{CP}^* - \hat{Y}_{IPPS}^2 + \hat{V}(\hat{Y}_{IPPS}) \right]^{\frac{1}{2}}$$

ผลที่ได้จากการวิจัยในครั้งนี้คือตัวประมาณค่าสัมประสิทธิ์การแปรผัน กรณีที่มีข้อมูลสูญหายที่เกิดจากการไม่ตอบเฉพาะบางคำถาม ผู้สนใจสามารถนำตัวประมาณค่าที่ผู้วิจัยได้นำเสนอไปใช้ในการเปรียบเทียบการกระจายของข้อมูล ที่มีประชากรตั้งแต่ 2 กลุ่มขึ้นไป โดยใช้ได้กับงานวิจัยที่มีการเก็บข้อมูลและมีข้อมูลบางส่วนที่ไม่สามารถเก็บรวบรวมข้อมูลได้ หรือกล่าวอีกนัยหนึ่งคือใช้ได้กับงานวิจัยหรือการศึกษาที่ต้องการเปรียบเทียบการกระจายของข้อมูล แต่มีข้อมูลบางส่วนสูญหาย นอกจากนี้ตัวประมาณค่าที่ผู้วิจัยนำเสนอเหมาะสมสำหรับกรณีที่ประชากรมีขนาดใหญ่ และตัวอย่างที่ถูกสุ่มมามีขนาดเล็ก (สัดส่วนตัวอย่างมีขนาดเล็ก) ทั้งนี้ตัวประมาณค่าที่ผู้วิจัยนำเสนอมีข้อจำกัดคือ ตัวประมาณค่าสัมประสิทธิ์การแปรผันที่ผู้วิจัยนำเสนอสามารถใช้เปรียบเทียบการกระจายเฉพาะข้อมูลของกลุ่มตัวอย่าง แต่ไม่สามารถใช้อ้างอิงถึงการกระจายของประชากรได้ เนื่องจากสูตรการคำนวณใช้เฉพาะข้อมูลตัวอย่างที่สุ่ม ดังนั้นผู้วิจัยมีข้อเสนอแนะที่ได้จากผลการวิจัยคือผู้ที่สนใจศึกษาสามารถขยายขอบเขตการวิจัยเพื่อหาตัวประมาณค่าความแปรปรวนของตัวประมาณค่าที่ผู้วิจัยนำเสนอ จากนั้นนำไปสร้างสถิติทดสอบสมมติฐานเพื่อใช้เปรียบเทียบการกระจายของข้อมูลระดับประชากรต่อไป

กิตติกรรมประกาศ

บทความวิจัยฉบับนี้ถูกล่วงไปด้วยดีด้วยความช่วยเหลือและการสนับสนุน เป็นอย่างยิ่งจากคณาจารย์ หลักสูตรสาขาวิชาคณิตศาสตร์และวิทยาการข้อมูล คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏเพชรบูรณ์

เอกสารอ้างอิง

- [1] Archana, V. & Aruna Rao, K. (2011). Improved Estimators of Coefficient of Variation in a Finite Population. *Statistics in Transition new series*, 12(2), 357-380.
- [2] Singh, R. & Mishra, M. (2019). Estimating Population Coefficient of Variation using a Single Auxiliary Variable in Simple Random Sampling. *Statistics in Transition new series*, 20(4), 89-111.
- [3] Das, A. K., & Tripathi, T. P. (1981). A class of estimators for co-efficient of variation using knowledge on coefficient of variation of an auxiliary character. In annual conference of Ind. Soc. Agricultural Statistics. Held at New Delhi, India.

- [4] Singh, H. P. & Solanki, R. S. (2012). An efficient class of estimators for the population mean using auxiliary information in systematic sampling, *Journal of Statistical Theory and Practice* 6(2), 274–285.
- [5] จินดา สวัสดิ์ทวี และสุชาดา กรเพชรปาณี. (2012). การประมาณค่าสัมประสิทธิ์การแปรผันแบบใหม่ของข้อมูลที่อยู่ในรูปอัตราส่วนที่สัมพันธ์กัน. *วารสารวิจัย มข.* 17 (2). 287-292.
- [6] Hansen, M.H. & Hurwitz, W.N. (1960). The problem of non-response in sample surveys. *Journal of the American Statistical Association*, 41(236), 517–529.
- [7] Archana, V. & Aruna Rao, K. (2011). Estimation of Co-efficient of Variation in PPS sampling , 58th Proceedings of the 2011 World Statistics Congress. 4019-4025.
- [8] ชูเกียรติ โพนแก้ว และหยาดพิรุณ ศุภรากรสกุล. (2559). การประมาณค่าความแปรปรวนของตัวประมาณค่าเฉลี่ยประชากรในกรณีที่มีข้อมูลไม่ตอบสนอง เมื่อสัดส่วนตัวอย่างมีขนาดเล็ก. การประชุมสัมมนาวิชาการและนำเสนอผลงานวิจัยระดับชาติ เครือข่ายบัณฑิตศึกษา มหาวิทยาลัยราชภัฏภาคเหนือ ครั้งที่ 16 และการประชุมวิชาการระดับชาติ มหาวิทยาลัยราชภัฏเพชรบูรณ์ ครั้งที่ 3 “งานวิจัยเพื่อพัฒนาท้องถิ่น”, มหาวิทยาลัยราชภัฏเพชรบูรณ์, 573-584.
- [9] Sampath, S. (2001). *Sampling theory and methods*. Norosa publishing, India.
- [10] วราฤทธิ์ พานิชกิจโกศลกุล. (2552). การจำลองแบบมอนติคาร์โลสำหรับประมาณค่าสัมประสิทธิ์ความแปรผันของการแจกแจงอินเวอร์สเกาส์เซียนเมื่อมีข้อมูลสูญหาย. *วารสาร มทร.อีสาน*, 2(2), 44-51.