



รายงานการวิจัย

ช่วงความเชื่อมั่นของค่าเฉลี่ยประชากรภายใต้การสุ่ม
โดยใช้ความน่าจะเป็นในการเลือกไม่เท่ากัน
**Confidence Interval for Population Mean
under Unequal Probability Sampling**

ชูเกียรติ โพนแก้ว

สาขาคณิตศาสตร์ คณะวิทยาศาสตร์และเทคโนโลยี

มหาวิทยาลัยราชภัฏเพชรบูรณ์

ประจำปีงบประมาณ 2562

รายงานวิจัยฉบับสมบูรณ์

ช่วงความเชื่อมั่นของค่าเฉลี่ยประชากรภายใต้การสุ่ม

โดยใช้ความน่าจะเป็นในการเลือกไม่เท่ากัน

Confidence Interval for Population Mean

under Unequal Probability Sampling

ชูเกียรติ โพนแก้ว

สาขาคณิตศาสตร์

คณะวิทยาศาสตร์และเทคโนโลยี

ทุนอุดหนุนโดย มหาวิทยาลัยราชภัฏเพชรบูรณ์

งบประมาณแผ่นดินที่พิจารณาโดยผ่านความเห็นชอบ

จากสำนักงานคณะกรรมการวิจัยแห่งชาติ

ประเภทโครงการสนับสนุนการวิจัยและนวัตกรรมเพื่อสร้างความเป็นเลิศทางวิชาการ

ประจำปีงบประมาณ 2562

(ก)

หัวข้อวิจัย ช่วงความเชื่อมั่นของค่าเฉลี่ยประชากรภายใต้การสุ่มโดยใช้ความน่าจะเป็นในการเลือกไม่เท่ากัน

ชื่อผู้วิจัย ชูเกียรติ โพนแก้ว

ชื่อหน่วยงาน สาขาวิชาคณิตศาสตร์ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏเพชรบูรณ์

ปีงบประมาณ 2562

บทคัดย่อ

การวิจัยมีวัตถุประสงค์ เพื่อนำเสนอ และเปรียบเทียบช่วงความเชื่อมั่นสำหรับประมาณค่าเฉลี่ยประชากร ภายใต้การสุ่มตัวอย่างแบบใช้ความน่าจะเป็นในการเลือกไม่เท่ากัน และไม่ได้สืบรวมถึงสัดส่วนตัวอย่างมีขนาดเล็ก ช่วงความเชื่อมั่นที่ผู้วิจัยนำเสนอประกอบด้วย ช่วงความเชื่อมั่นแบบเชิงเส้น และช่วงความเชื่อมั่นแบบไม่เป็นเชิงเส้น ผู้วิจัยเปรียบเทียบประสิทธิภาพของช่วงความเชื่อมั่นที่นำเสนอ โดยใช้เทคนิคมอนติคาร์โลซิมูเลชัน และใช้ขนาดตัวอย่างทั้งหมด 5 ระดับ ประกอบด้วย 30, 35, 40, 45, และ 50 เกณฑ์ที่ใช้ในการพิจารณาประสิทธิภาพช่วงความเชื่อมั่นที่นำเสนอได้แก่ ความกว้างเฉลี่ยของช่วงความเชื่อมั่นที่ระดับความเชื่อมั่น เท่ากับ 90% 95% และ 99% ผลการเปรียบเทียบประสิทธิภาพพบว่าช่วงความเชื่อมั่นแบบไม่เป็นเชิงเส้น มีประสิทธิภาพดีกว่าช่วงความเชื่อมั่นที่ไม่เป็นเชิงเส้น

คำสำคัญ : ตัวประมาณค่า, สัดส่วนตัวอย่าง, ความแปรปรวน

Research 's Title	Confidence Interval for Population Mean under Unequal Probability Sampling
Researcher	Chugiat Ponkaew
Department	Mathematics
	Phetchabun Rajabhat University Year 2019

Abstract

In this research paper, we aim to propose and compare confidence interval estimators for estimating population mean under unequal probability sampling without replacement and overall negligible sampling fraction. The propose estimators comprise linear and ratio confidence interval. We compare the efficiency of the propose estimators by using mote carlo simulation with 5 levels sample size there are 30, 35, 40, 45 and 50. The considered criterions are based on expected length of confidence interval with 90%, 95% and 99% confidence levels. The simulation study show that for all sample size and confidence levels the nonlinear confidence interval is more efficiency than linear confidence interval.

Keywords : Estimator, Sampling fraction, Variance

สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	ก
บทคัดย่อภาษาอังกฤษ.....	ข
กิตติกรรมประกาศ.....	ค
สารบัญ.....	ง
สารบัญตาราง.....	ฉ
สารบัญรูป.....	ช
บทที่ 1 บทนำ.....	1
1.1 ที่มาและความสำคัญของปัญหา.....	1
1.2 วัตถุประสงค์.....	5
1.3 สมมติฐานการวิจัย.....	5
1.4 ประโยชน์ที่คาดว่าจะได้รับ.....	5
1.5 ขอบเขตการวิจัย.....	5
1.6 นิยามศัพท์ที่เกี่ยวข้อง.....	6
1.7 กรอบแนวคิดในการวิจัย.....	6
บทที่ 2 เอกสารและงานวิจัยที่เกี่ยวข้อง.....	8
2.1 การประมาณค่า.....	8
2.2 เกณฑ์ที่ใช้ในการเปรียบเทียบประสิทธิภาพช่วงความเชื่อมั่น.....	19
2.3 การหาความแปรปรวนของตัวประมาณค่าแบบไม่เป็นเชิงเส้น.....	20
2.4 งานวิจัยที่เกี่ยวข้อง.....	22
บทที่ 3 วิธีดำเนินการวิจัย.....	24
3.1 สัญลักษณ์และข้อตกลงเบื้องต้น.....	24
3.2 การนำเสนอช่วงความเชื่อมั่นสำหรับการประมาณค่าเฉลี่ยประชากร.....	25
3.3 การเปรียบเทียบประสิทธิภาพช่วงความเชื่อมั่นที่นำเสนอ.....	25
3.4 เกณฑ์การเปรียบเทียบประสิทธิภาพช่วงความเชื่อมั่น.....	26

สารบัญ (ต่อ)

	หน้า
บทที่ 4 ผลการวิจัย.....	28
4.1 ช่วงความเชื่อมั่นสำหรับประมาณค่าเฉลี่ยประชากรที่น่าเสนอ.....	28
4.2 ผลการเปรียบเทียบช่วงความเชื่อมั่นสำหรับประมาณค่าเฉลี่ยประชากรที่ นำเสนอ.....	32
บทที่ 5 สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ.....	34
5.1 สรุปผลการวิจัย.....	34
5.2 การอภิปรายผล.....	35
5.3 ข้อเสนอแนะ.....	35
บรรณานุกรม.....	36
ภาคผนวก.....	46
ประวัติผู้วิจัย.....	60

สารบัญตาราง

		หน้า
ตารางที่		
4.1	ค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่นแบบเชิงเส้น และช่วงความเชื่อมั่นแบบไม่เป็นเชิงเส้น ที่ระดับความเชื่อมั่น 90%.....	32
4.2	ค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่นแบบเชิงเส้น และช่วงความเชื่อมั่นแบบไม่เป็นเชิงเส้น ที่ระดับความเชื่อมั่น 95%.....	33
4.3	ค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่นแบบเชิงเส้น และช่วงความเชื่อมั่นแบบไม่เป็นเชิงเส้น ที่ระดับความเชื่อมั่น 99%.....	33

(ข)

สารบัญรูป

รูปที่	หน้า
1.1	กรอบแนวคิดในการทำวิจัย..... 7

(ก)

กิตติกรรมประกาศ

ขอขอบคุณสถาบันวิจัยและพัฒนา มหาวิทยาลัยราชภัฏเพชรบูรณ์ ที่ให้ทุนสนับสนุนประจำปี
งบประมาณ 2562

ขอขอบพระคุณ ผู้ช่วยศาสตราจารย์อาดุลย์ จงรักษ์ และอาจารย์ หยาตพิรุณ สุภรากรสกุล
ที่ได้ให้ความอนุเคราะห์เป็นผู้ทรงคุณวุฒิตรวจสอบเครื่องมือ และเนื้อหาในโครงการวิจัย

ท้ายที่สุดนี้ ขอขอบคุณผู้ให้ความช่วยเหลือและคำแนะนำจากผู้ที่เกี่ยวข้องทุกท่าน ตลอดจน
บิดา มารดาและสมาชิกในครอบครัวทุกคนที่คอยเป็นกำลังใจ ให้การสนับสนุนทุกอย่างที่เกี่ยวข้อง
กับงานวิจัยทั้งทางตรงและทางอ้อม จนงานวิจัยนี้สำเร็จลงได้

ชูเกียรติ โพนแก้ว

30 กรกฎาคม 2562

บทที่ 1

บทนำ

1.1 ที่มาและความสำคัญของปัญหา

การประมาณค่าช่วงความเชื่อมั่น (Confidence Interval) เป็นวิธีการทางสถิติที่สำคัญที่ใช้ในการประมาณพารามิเตอร์ (Parameter) ที่สนใจซึ่งเป็นลักษณะเฉพาะของประชากร ค่าเฉลี่ยเป็นหนึ่งในพารามิเตอร์ ที่นักสถิติ หรือนักวิจัยสนใจศึกษาและต้องการทราบค่า โดยปกติค่าเฉลี่ยของประชากรคำนวณได้จากผลรวมของคุณลักษณะที่สนใจทุกหน่วยตัวอย่างของประชากรหารด้วยขนาดของประชากร แต่ในบางกรณีเช่นประชากรมีขนาดใหญ่ หรือไม่ทราบขนาดที่แน่นอนของประชากร ค่าเฉลี่ยประชากรไม่สามารถคำนวณค่าได้ ดังนั้นจึงจำเป็นต้องประมาณค่าดังกล่าวจากข้อมูลของตัวอย่าง และเรียกวิธีการดังกล่าวว่าการประมาณค่า (Estimation)

การประมาณค่าพารามิเตอร์ แบ่งออกเป็นการประมาณค่าแบบจุด (Point estimation) และการประมาณค่าแบบช่วง (Interval estimation) การประมาณค่าแบบจุดเป็นการประมาณค่าพารามิเตอร์ที่สนใจเพียงค่าเดียว จากข้อมูลตัวอย่าง ทำให้ค่าประมาณที่ได้อาจจะมียกเว้นค่าพารามิเตอร์ หรือมีค่าที่คลาดเคลื่อนไปจากค่าพารามิเตอร์ขึ้นอยู่กับตัวอย่างที่ถูกสุ่มจากประชากร ส่วนการประมาณค่าแบบช่วงเป็นการประมาณค่าพารามิเตอร์ของประชากร โดยใช้ข้อมูลของตัวอย่างสร้างตัวสถิติที่ครอบคลุมพารามิเตอร์ขึ้นมาสองค่า และเรียกตัวสถิติดังกล่าวว่า จีความเชื่อมั่นล่าง และจีความเชื่อมั่นบน

พิจารณาประชากร $U = \{1, 2, \dots, N\}$ ขนาดจำกัด ขนาด N ตัวแปรที่สนใจศึกษา (Study variable) แทนด้วยตัวแปร y และพารามิเตอร์ที่ต้องการประมาณค่าคือค่าเฉลี่ยของประชากร เขียนแทนด้วยสัญลักษณ์ $\bar{Y} = N^{-1} \sum_{i \in U} y_i$ ซึ่ง y_i แทนค่าสังเกตของประชากรหน่วยที่ i เมื่อ $i \in U$ กรณีที่ตัวอย่าง s ขนาด n ถูกสุ่มโดยใช้แผนการสุ่มตัวอย่างแบบง่าย (Simple random sampling) ตัวประมาณค่าเฉลี่ยประชากรแบบจุดเขียนแทนด้วย $\hat{\bar{Y}}$ และกำหนดค่าตามที่แสดงในสมการที่ (1-1)

$$\hat{\bar{Y}} = n^{-1} \sum_{i \in s} y_i \quad (1-1)$$

ภายใต้การสุ่มแบบง่ายและไม่ใส่คืน (Simple random sampling without replacement) หรือแผนการสุ่มตัวอย่างแบบ SRSWOR ค่าความแปรปรวนของ $\hat{\bar{Y}}$ เขียนแทนด้วย $V(\hat{\bar{Y}})$ และกำหนดค่าตามที่แสดงในสมการที่ (1-2)

$$V(\hat{Y}) = (1-f) \frac{S^2}{n} \quad (1-2)$$

เมื่อ $f = nN^{-1}$ แทนสัดส่วนตัวอย่าง (Sampling fraction) และ $S^2 = (N-1)^{-1} \sum_{i \in U} (y_i - \bar{Y})^2$ โดยปกติค่า $V(\hat{Y})$ จะไม่ทราบค่าเนื่องจากค่า S^2 ขึ้นอยู่กับข้อมูลของประชากรทำให้ไม่สามารถคำนวณค่าได้ ดังนั้นจำเป็นต้องประมาณค่า $V(\hat{Y})$ จากข้อมูลของตัวอย่าง ซึ่งตัวประมาณค่าของ $V(\hat{Y})$ เขียนแทนด้วยสัญลักษณ์ $\hat{V}(\hat{Y})$ และกำหนดค่าตามที่แสดงในสมการที่ (1-3)

$$\hat{V}(\hat{Y}) = (1-f) \frac{\hat{S}^2}{n} \quad (1-3)$$

เมื่อ $\hat{S}^2 = (n-1)^{-1} \sum_{i \in s} (y_i - \hat{Y})^2$ และ $\hat{Y}$ กำหนดค่าตามที่แสดงในสมการ (1-1) กรณีที่ตัวอย่างมีขนาดใหญ่ ($n \geq 30$) กำหนดให้ Z เป็นตัวแปรสุ่มและมีค่าตามที่แสดงในสมการ (1-4)

$$Z = \frac{\hat{Y} - E(\hat{Y})}{\sqrt{\hat{V}(\hat{Y})}} = \frac{\hat{Y} - \bar{Y}}{\sqrt{\hat{V}(\hat{Y})}} \quad (1-4)$$

เมื่อใช้ทฤษฎีลิมิตเข้าสู่ส่วนกลาง (Central limit theorem) สามารถแสดงได้ว่า $Z = \frac{\hat{Y} - \bar{Y}}{\sqrt{\hat{V}(\hat{Y})}} \sim \text{app}N(0,1)$ จากนั้นหาช่วงความเชื่อมั่นตามวิธีของ [กัลยา วาณิชย์บัญชา \(2547\)](#) และ [Casella และ Berger \(2002\)](#) จะได้ว่าช่วงความเชื่อมั่นของ $\bar{Y}$ ที่ระดับ $(1-\alpha)100\%$ เขียนแทนด้วยสัญลักษณ์ $C.I._{(0)}$ มีค่าตามที่แสดงในสมการ (1-5)

$$C.I._{(0)} = \left[\hat{Y} - z_{1-\frac{\alpha}{2}} \sqrt{\hat{V}(\hat{Y})}, \hat{Y} + z_{1-\frac{\alpha}{2}} \sqrt{\hat{V}(\hat{Y})} \right] \quad (1-5)$$

เมื่อ $z_{1-\frac{\alpha}{2}}$ แทนค่าของตัวแปรสุ่มแบบปกติมาตรฐาน Z ที่ทำให้ $P(Z < z_{1-\alpha/2}) = 1-\alpha/2$ และ α แทนความน่าจะเป็นที่ช่วงความเชื่อมั่นที่สร้างขึ้นไม่ครอบคลุม $\bar{Y}$

การประมาณค่า $\bar{Y}$ แบบจุดตามที่แสดงในสมการที่ (1-1) และการประมาณค่าแบบช่วงตามที่แสดงในสมการที่ (1-5) ศึกษาภายใต้การสุ่มแบบ SRSWOR ซึ่งเป็นการสุ่มโดยเลือกหน่วยตัวอย่างด้วยความน่าจะเป็นเท่ากัน (Equal probability sampling) ทุกหน่วยตัวอย่าง อย่างไรก็ตามในกรณีที่ทราบข้อมูลของตัวแปรช่วย (Auxiliary variable) ที่มีความสัมพันธ์กับตัวแปรที่สนใจศึกษา เขียนแทนด้วยสัญลักษณ์ k และทราบค่าทุกหน่วยตัวอย่างหรือทราบค่า k_i ทุก $i \in U$ ตัวประมาณค่าของ $\bar{Y}$ ทั้งตัวประมาณค่าแบบจุด และตัวประมาณค่าแบบช่วงสามารถเพิ่มประสิทธิภาพได้โดยใช้การสุ่มตัวอย่างโดยใช้ความน่าจะเป็นในการเลือกไม่เท่ากัน (Unequal probability sampling) ซึ่ง

ภายใต้แผนการสุ่มดังกล่าวและเป็นแบบใส่คืน Hansen และ Horvitz (1943) ได้นำเสนอตัวประมาณค่าเฉลี่ยประชากรและมีค่าตามที่แสดงในสมการที่ (1-6)

$$\hat{\bar{Y}}_{HH} = N^{-1} \sum_{i \in S} f(y_i) = N^{-1} \sum_{i \in S} b_i \quad (1-6)$$

เมื่อ $b_i = f(y_i)$ ส่วนตัวประมาณค่าความแปรปรวนของ $\hat{\bar{Y}}_{HH}$ หรือ $\hat{V}(\hat{\bar{Y}}_{HH})$ มีค่าตามที่แสดงในสมการที่ (1-7)

$$\hat{V}(\hat{\bar{Y}}_{HH}) = \frac{1}{N^2} \left(\frac{n}{n-1} \right) \left[\sum_{i \in S} \hat{b}_i^2 - \frac{1}{n} \left(\sum_{i \in S} \hat{b}_i \right)^2 \right] \quad (1-7)$$

เมื่อ $\hat{b}_i$ แทนตัวประมาณค่าของ b_i ถ้า b_i มีพารามิเตอร์ที่ไม่ทราบค่า และ $\hat{b}_i = b_i$ ในกรณีอื่นๆ

ต่อมา Horvitz และ Thompson (1952) ได้ขยายตัวประมาณค่าของ Hansen และ Horvitz (1943) ไปยังแผนการสุ่มตัวอย่างแบบใช้ความน่าจะเป็นในการเลือกไม่เท่ากันและไม่ใส่คืน ซึ่งตัวประมาณค่าที่ Horvitz และ Thompson (1952) นำเสนออยู่ในรูปตัวประมาณค่าแบบเชิงเส้น เขียนแทนด้วย $\hat{\bar{Y}}_{HT}$ และมีค่าตามที่แสดงในสมการที่ (1-8)

$$\hat{\bar{Y}}_{HT} = N^{-1} \sum_{i \in S} \frac{y_i}{\pi_i} \quad (1-8)$$

สำหรับตัวประมาณค่าความแปรปรวนของ $\hat{\bar{Y}}_{HT}$ เขียนแทนด้วย $\hat{V}(\hat{\bar{Y}}_{HT})$ และมีค่าตามที่แสดงในสมการที่ (1-9)

$$\hat{V}(\hat{\bar{Y}}_{HT}) = \sum_{i \in S} \left(\frac{1 - \pi_i}{\pi_i^2} \right) y_i^2 + \sum_{i \in S} \sum_{j \in S, j \neq i} \left(\frac{\pi_{ij} - \pi_i \pi_j}{\pi_{ij} \pi_i \pi_j} \right) y_i y_j \quad (1-9)$$

เมื่อ π_i แทนความน่าจะเป็นที่หน่วยตัวอย่างที่ i ตกอยู่ในตัวอย่าง s ขนาด n หรือ $\pi_i = P(i \in s)$ ส่วนค่า π_{ij} แทนความน่าจะเป็นที่หน่วยตัวอย่างที่ i และ j ตกอยู่ในตัวอย่าง s ขนาด n หรือ $\pi_{ij} = P(i \wedge j \in s)$ เมื่อ $i, j \in U$ กรณีที่ตัวอย่างมีขนาดใหญ่ และใช้ทฤษฎีลิมิตเข้าสู่ส่วนกลาง (Central limit theorem) สามารถแสดงได้ว่า $Z = \frac{\hat{\bar{Y}}_{HT} - \bar{Y}}{\sqrt{\hat{V}(\hat{\bar{Y}}_{HT})}} \sim \text{app}N(0,1)$ และช่วงความเชื่อมั่น $(1-\alpha)100\%$

ของ $\bar{Y}$ เขียนแทนด้วย $C.I._{(1)}$ และกำหนดค่าตามที่แสดงในสมการที่ (1-10)

$$C.I._{(1)} = \left[\hat{\bar{Y}}_{HT} - z_{1-\frac{\alpha}{2}} \sqrt{\hat{V}(\hat{\bar{Y}}_{HT})}, \hat{\bar{Y}}_{HT} + z_{1-\frac{\alpha}{2}} \sqrt{\hat{V}(\hat{\bar{Y}}_{HT})} \right] \quad (1-10)$$

เมื่อ $z_{1-\frac{\alpha}{2}}$ กำหนดค่าตามที่แสดงในสมการที่ (1-5) ต่อมา Hájek (1971) ได้นำเสนอตัวประมาณค่าเฉลี่ยประชากรแบบไม่เป็นเชิงเส้น (Nonlinear estimator) และกำหนดค่าตามที่แสดงในสมการที่ (1-11)

$$\hat{Y}_{Hajek} = \sum_{i \in S} \frac{y_i}{\pi_i} / \sum_{i \in S} \frac{1}{\pi_i} \quad (1-11)$$

Särndal , Swenson และ Wretman (1992) กล่าวว่ากรณีที่ (1) y_i มีความสัมพันธ์ในทิศทางตรงข้ามกับ π_i หรือ (2) $y_i - \bar{Y}$ มีขนาดเล็กทุกค่า $i \in U$ ตัวประมาณค่าที่ Hájek (1971) นำเสนอ จะเป็นตัวประมาณค่าที่มีประสิทธิภาพดีกว่าตัวประมาณของ Horvitz และ Thompson (1952) ต่อมา Tillé และ Matei (2016) ได้ทำการเปรียบเทียบประสิทธิภาพของ $\hat{Y}_{HT}$ และ $\hat{Y}_{Hajek}$ ภายใต้เงื่อนไขที่ Särndal , Swenson และ Wretman (1992) กล่าวถึง โดยใช้เทคนิคมอนติคาร์โลซิมูเลชัน และเปรียบเทียบผลจากค่าความคลาดเคลื่อนกำลังสอง (Mean square error) และพบว่าสอดคล้องกับคำว่าของ Särndal , Swenson และ Wretman (1992)

จากสมการที่ (1-11) จะเห็นว่า $\hat{Y}_{Ha'jek}$ อยู่ในรูปของตัวประมาณค่าแบบไม่เป็นเชิงเส้น ดังนั้นจำเป็นต้องใช้เทเลอร์แบบเชิงเส้นในการปรับตัวประมาณค่าดังกล่าวให้อยู่ในรูปแบบเชิงเส้นก่อนการหาตัวประมาณค่าความแปรปรวน และมีค่าตามที่แสดงในสมการที่ (1-12)

$$\hat{V}(\hat{Y}_{Ha'jek}) \approx \frac{1}{N^2} \left[\sum_{i \in S} \frac{1-\pi_i}{\pi_i^2} (y_i - \hat{Y}_{Ha'jek})^2 + \sum_{i \in S} \sum_{j \in S, j \neq i} \frac{\pi_{ij} - \pi_i \pi_j}{\pi_{ij} \pi_i \pi_j} (y_i - \hat{Y}_{Ha'jek})(y_j - \hat{Y}_{Ha'jek}) \right] \quad (1-12)$$

กรณีที่ $n \geq 30$ และใช้ทฤษฎีลิมิตเข้าสู่ส่วนกลาง (Central limit theorem) สามารถแสดงได้ว่า $Z = \frac{\hat{Y}_{Ha'jek} - \bar{Y}}{\sqrt{\hat{V}(\hat{Y}_{Ha'jek})}} \sim appN(0,1)$ และช่วงความเชื่อมั่น $(1-\alpha)100\%$ ของ $\bar{Y}$ เขียนแทนด้วย $C.I._{(2)}$

และกำหนดค่าตามที่แสดงในสมการที่ (1-13)

$$C.I._{(2)}(\bar{Y}) = \left[\hat{Y}_{Ha'jek} - z_{1-\frac{\alpha}{2}} \sqrt{\hat{V}(\hat{Y}_{Ha'jek})}, \hat{Y}_{Ha'jek} + z_{1-\frac{\alpha}{2}} \sqrt{\hat{V}(\hat{Y}_{Ha'jek})} \right] \quad (1-13)$$

เมื่อ $\hat{Y}_{Ha'jek}$ และ $\hat{V}(\hat{Y}_{Ha'jek})$ กำหนดค่าตามที่แสดงในสมการ (1-11) และ (1-12) ตามลำดับ

จากสมการที่ (1-10) และสมการที่ (1-13) จะเห็นว่าช่วงความเชื่อมั่นของ $\bar{Y}$ ภายใต้การสุ่มแบบใช้ความน่าจะเป็นในการเลือกไม่เท่ากันและไม่ใส่คืนที่สร้างจากตัวประมาณค่าของ Horvitz และ Thompson (1952) และตัวประมาณค่าของ Hájek (1971) ก่อนข้างมีความยุ่งยากในการคำนวณค่าเนื่องจากการคำนวณค่า $\hat{V}(\hat{Y}_{HT})$ และ $\hat{V}(\hat{Y}_{Ha'jek})$ ต้องการค่า π_{ij} ซึ่งปกติจะมีความซับซ้อนในการกำหนดค่าที่แท้จริง (Berger, 2004) อย่างไรก็ตามกรณีที่สัดส่วนตัวอย่าง ($f = nN^{-1}$) มีขนาด

เล็ก (ขนาดตัวอย่างน้อยกว่าหรือเท่ากับ 5% ของขนาดประชากร) ตัวประมาณค่าความแปรปรวนของ ตัวประมาณค่าเฉลี่ยประชากรภายใต้แผนการสุ่มตัวอย่างแบบใช้ความน่าจะเป็นในการเลือกไม่เท่ากันและไม่ได้คืน สามารถประมาณค่าด้วยตัวประมาณค่าที่ Hansen และ Horvitz (1943) นำเสนอตามที่แสดงในสมการที่ (1-7) ดังนั้นการวิจัยในครั้งนี้ผู้วิจัยต้องการพัฒนาช่วงความเชื่อมั่นสำหรับประมาณค่าเฉลี่ยประชากร ที่สร้างตัวประมาณค่าของ Horvitz และ Thompson (1952) และตัวประมาณค่าของ Hájek (1971) ตามที่แสดงในสมการที่ (1-10) และสมการที่ (1-13) ด้วยการประมาณค่าความแปรปรวนของ $\hat{Y}_{HT}$ และ $\hat{Y}_{Hajek}$ ด้วยวิธีการของ Hansen และ Horvitz (1943) และเรียกช่วงความเชื่อมั่นที่ได้ว่าช่วงความเชื่อมั่นแบบเชิงเส้น และช่วงความเชื่อมั่นแบบไม่เป็นเชิงเส้นตามลำดับ

1.2 วัตถุประสงค์

สำหรับงานวิจัยในครั้งนี้มีวัตถุประสงค์เพื่อ

1.2.1 นำเสนอช่วงความเชื่อมั่นของค่าเฉลี่ยประชากร ภายใต้การสุ่มตัวอย่างโดยใช้ความน่าจะเป็นในการเลือกไม่เท่ากัน และไม่ได้คืน

1.2.2 เปรียบเทียบประสิทธิภาพของช่วงความเชื่อมั่นที่นำเสนอ โดยพิจารณาจากค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่น

1.3 สมมติฐานการวิจัย

การวิจัยในครั้งนี้กำหนดสมมติฐานคือ ช่วงความเชื่อมั่นแบบไม่เป็นเชิงเส้น มีประสิทธิภาพดีกว่าช่วงความเชื่อมั่นแบบเชิงเส้น

1.4 ประโยชน์ที่คาดว่าจะได้รับ

ผลการวิจัยทำให้เกิดประโยชน์ด้านวิชาการ เนื่องจากการวิจัยในครั้งนี้ทำการเพิ่มประสิทธิภาพในการประมาณค่าเฉลี่ยประชากร โดยใช้ตัวประมาณค่าแบบช่วง และสุ่มตัวอย่างโดยใช้ความน่าจะเป็นในการเลือกไม่เท่ากัน ซึ่งเป็นประโยชน์สำหรับหน่วยงานที่ทำการสำรวจ และต้องการประมาณค่าเฉลี่ยของประชากร เช่น สถิติจังหวัด สำนักงานขนส่งจังหวัด รวมถึงเป็นประโยชน์แก่นักศึกษามหาวิทยาลัยราชภัฏเพชรบูรณ์ ที่ลงทะเบียนเรียนรายวิชาทางสถิติ

1.5 ขอบเขตของการวิจัย

1.5.1 กำหนดให้ y แทนตัวแปรที่สนใจศึกษา และไม่มีการสูญหายของข้อมูล

1.5.2 สมมติว่าตัวอย่าง s ขนาด n ถูกสุ่มจากประชากร U ขนาด N ภายใต้แผนการสุ่มตัวอย่างแบบใช้ความน่าจะเป็นในการเลือกไม่เท่ากันและไม่ใส่คืน

1.5.3 ตัวแปรกำหนดขนาด k มีความสัมพันธ์กับตัวแปร y และทราบค่าทุกค่า k_i เมื่อ $i \in U$

1.6 นิยามศัพท์ที่เกี่ยวข้อง

ตัวประมาณค่าแบบเชิงเส้น หมายถึง ตัวประมาณค่าเฉลี่ยประชากรของ Horvitz และ Thompson (1952)

ตัวประมาณค่าแบบไม่เป็นเชิงเส้น หมายถึง ตัวประมาณค่าเฉลี่ยประชากรของ Hájek (1971)

ตัวประมาณค่าความแปรปรวนแบบเชิงเส้น หมายถึง ตัวประมาณค่าความแปรปรวนของตัวประมาณค่าแบบเชิงเส้น

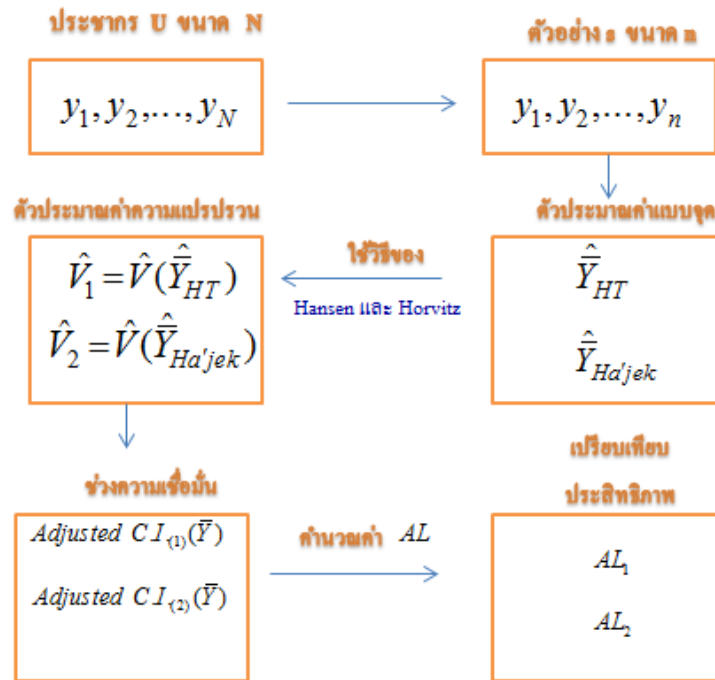
ตัวประมาณค่าความแปรปรวนแบบไม่เป็นเชิงเส้น หมายถึง ตัวประมาณค่าความแปรปรวนของตัวประมาณค่าแบบไม่เป็นเชิงเส้น

ช่วงความเชื่อมั่นแบบเชิงเส้น หมายถึง ช่วงความเชื่อมั่นของค่าเฉลี่ยประชากรที่สร้างจากตัวประมาณค่าแบบเชิงเส้น และตัวประมาณค่าความแปรปรวนแบบเชิงเส้น

ช่วงความเชื่อมั่นแบบไม่เป็นเชิงเส้น หมายถึง ช่วงความเชื่อมั่นของค่าเฉลี่ยประชากรที่สร้างจาก ตัวประมาณค่าแบบไม่เป็นเชิงเส้น และตัวประมาณค่าความแปรปรวนแบบไม่เป็นเชิงเส้น

1.7 กรอบแนวคิดในการทำวิจัย

การศึกษาช่วงความเชื่อมั่นของค่าเฉลี่ยประชากร ภายใต้การสุ่มโดยใช้ความน่าจะเป็นในการเลือกไม่เท่ากัน มีกรอบแนวคิดในการทำวิจัยตามที่แสดงในภาพที่ 1.1



ภาพที่ 1.1 กรอบแนวคิดในการวิจัย

จากภาพที่ 1.1 จากประชากร U ขนาด N สุ่มตัวอย่าง s ขนาด n โดยใช้แผนการสุ่มตัวอย่างแบบใช้ความน่าจะเป็นในการเลือกไม่เท่ากันและไม่ใส่คืน คำนวณค่า $\hat{Y}_{HT}$ และ $\hat{Y}_{Ha'jek}$ จากตัวอย่าง s ขนาด n ที่สุ่มได้ เมื่อ $\hat{Y}_{HT}$ และ $\hat{Y}_{Ha'jek}$ แทนตัวประมาณค่าแบบเชิงเส้น และตัวประมาณค่าแบบไม่เป็นเชิงเส้นตามลำดับ จากนั้นหาตัวประมาณค่าความแปรปรวนของ $\hat{Y}_{HT}$ และ $\hat{Y}_{Ha'jek}$ ด้วยวิธีของ Hansen และ Horvitz (1943) และเรียกตัวประมาณค่าที่ได้ว่าตัวประมาณค่าความแปรปรวนแบบเชิงเส้น เขียนแทนด้วย $\hat{V}_1$ และตัวประมาณค่าความแปรปรวนแบบไม่เป็นเชิงเส้น เขียนแทนด้วย $\hat{V}_2$ ตามลำดับ ต่อมาใช้ $\hat{Y}_{HT}$ และ $\hat{V}_1$ สร้างความเชื่อมั่นที่เป็นเชิงเส้น ตามวิธีของกัลยา วานิชย์บัญชา (2547) และ Casella และ Berger (2002) เขียนแทนด้วย $Adjusted\ C.I_{(1)}(\bar{Y})$ ในทำนองเดียวกันเมื่อใช้วิธีของกัลยา วานิชย์บัญชา (2547) และ Casella และ Berger (2002) สร้างช่วงความเชื่อมั่นที่ไม่เป็นเชิงเส้นจาก $\hat{Y}_{Ha'jek}$ และ $\hat{V}_2$ เขียนแทนด้วยสัญลักษณ์ $Adjusted\ C.I_{(2)}(\bar{Y})$

หลังจากนำเสนอช่วงความเชื่อมั่นที่เป็นเชิงเส้น และช่วงความเชื่อมั่นที่ไม่เป็นเชิงเส้น ผู้วิจัยทำการเปรียบเทียบประสิทธิภาพของช่วงความเชื่อมั่นที่นำเสนอ โดยใช้เทคนิคมอนติคาร์โล ซิมูเลชัน โดยพิจารณาจากค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่นของ $Adjusted\ C.I_{(1)}(\bar{Y})$ และ $Adjusted\ C.I_{(2)}(\bar{Y})$

บทที่ 2

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

บทนี้ผู้วิจัยจะได้ถึงทฤษฎีพื้นฐานเกี่ยวกับช่วงความเชื่อมั่นของค่าเฉลี่ยประชากร 1 กลุ่ม กรณีที่ตัวอย่างมีขนาดใหญ่ การเปรียบเทียบประสิทธิภาพช่วงความเชื่อมั่นด้วยความกว้างเฉลี่ยของความกว้างของช่วงความเชื่อมั่น (Average length of the confidence interval) และหัวข้อสุดท้ายเป็นการรวบรวมทฤษฎี และงานวิจัยที่เกี่ยวข้อง

2.1 การประมาณค่า

2.1.1 ความหมายของการประมาณค่า

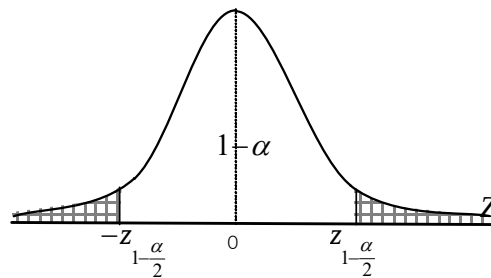
การประมาณค่าโดยใช้ช่วงความเชื่อมั่น (Confidence Interval) เป็นวิธีการทางสถิติที่สำคัญเพื่อใช้ในการประมาณพารามิเตอร์ (Parameter) ที่สนใจซึ่งเป็นลักษณะเฉพาะของประชากร ค่าเฉลี่ยเป็นหนึ่งในพารามิเตอร์ ที่นักสถิติ หรือนักวิจัยสนใจศึกษาและต้องการทราบค่า โดยปกติค่าเฉลี่ยของประชากรคำนวณได้จากผลรวมของคุณลักษณะที่สนใจทุกหน่วยตัวอย่างของประชากรหารด้วยขนาดของประชากร แต่ในบางกรณีเช่นประชากรมีขนาดใหญ่ หรือไม่ทราบขนาดที่แท้จริงของประชากร ค่าเฉลี่ยประชากรไม่สามารถคำนวณค่าได้ ดังนั้นจึงจำเป็นต้องประมาณค่าดังกล่าวด้วยตัวสถิติ (Statistic) ที่คำนวณได้จากตัวอย่าง (กัลยา วานิชย์บัญชา, 2547) การประมาณค่าแบ่งออกเป็นการประมาณค่าแบบจุด และการประมาณค่าแบบช่วง การประมาณค่าแบบจุดเป็นการประมาณค่าพารามิเตอร์ที่สนใจเพียงค่าเดียว โดยใช้ข้อมูลของตัวอย่าง ซึ่งค่าประมาณที่ได้อาจจะมีความคลาดเคลื่อนไปจากค่าพารามิเตอร์ที่สนใจอยู่บ้าง ตัวอย่างที่ถูกสุ่มจากประชากร ส่วนการประมาณค่าแบบช่วงเป็นการประมาณค่าพารามิเตอร์ของประชากร โดยใช้ข้อมูลของตัวอย่างสร้างตัวสถิติที่ครอบคลุมพารามิเตอร์ขึ้นมาสองค่า และเรียกตัวสถิติดังกล่าวว่าขีดความเชื่อมั่นล่าง และขีดความเชื่อมั่นบน เขียนแทนด้วยสัญลักษณ์ L และ U ตามลำดับ การประมาณค่าแบบช่วงจำเป็นต้องกำหนดระดับความเชื่อมั่น (Confidence Interval) เขียนแทนด้วย $1-\alpha$ เพื่อบอกถึงโอกาสที่พารามิเตอร์ของประชากร จะมีค่าอยู่ระหว่าง L และ U เช่น $P(L < \theta < U) = 0.95$ เมื่อ θ แทนพารามิเตอร์ที่สนใจ หมายถึงโอกาสที่ θ จะมีค่าอยู่ระหว่าง L และ U เป็น 0.95 หรือ 95% และอีก 0.05 หรือ 5% มีโอกาสที่ $\theta < L$ หรือ $\theta > U$ หรือกล่าวได้ว่า ในการสุ่มตัวอย่าง s ขนาด n ทั้งหมด 100 เพื่อสร้างตัวประมาณค่าแบบช่วง จะมีเพียง 5 ครั้งที่ θ อยู่นอกช่วงที่สร้างขึ้น(กัลยา วานิชย์บัญชา, 2547)

2.1.2 ช่วงความเชื่อมั่นสำหรับประมาณค่าเฉลี่ยประชากร 1 กลุ่ม

พิจารณาประชากร $U=\{1,2,\dots,N\}$ ขนาดจำกัด ขนาด N ตัวแปรที่สนใจศึกษาแทนด้วยตัวแปร y และพารามิเตอร์ที่ต้องการประมาณค่าคือค่าเฉลี่ยของประชากร เขียนแทนด้วยสัญลักษณ์ $\bar{Y}=N^{-1}\sum_{i\in U}y_i$ ซึ่ง y_i แทนค่าสังเกตของประชากรหน่วยที่ i เมื่อ $i\in U$ การประมาณค่าเฉลี่ยสำหรับประชากรหนึ่งกลุ่ม เมื่อต้องการประมาณค่าแบบจุดสามารถใช้ค่าเฉลี่ยของตัวอย่าง $\hat{Y}$ เป็นตัวประมาณ แต่เมื่อต้องการประมาณค่าเฉลี่ยประชากรแบบช่วง จำเป็นต้องทราบการแจกแจงและความแปรปรวนของ $\hat{Y}$ เพื่อสร้างความเชื่อมั่น กรณีที่ไม่ทราบการแจกแจงของ $\hat{Y}$ (ไม่ทราบการแจกแจงของตัวสถิติ) จำเป็นต้องสุ่มตัวอย่างให้มีขนาดใหญ่ ซึ่งโดยปกติจะสุ่มตัวอย่างให้มีขนาดมากกว่าหรือเท่ากับ 30 ($n\geq 30$) เพื่อให้ทฤษฎีลิมิตเข้าสู่ส่วนกลาง ประมาณการแจกแจงของตัวสถิติและสร้างความเชื่อมั่นดังนี้ กำหนดให้ $E(\hat{Y})=\bar{Y}$ และ $V(\hat{Y})$ แทนค่าคาดหมาย และความแปรปรวนของ $\hat{Y}$ และ $\hat{V}(\hat{Y})$ แทนตัวประมาณค่าของ $V(\hat{Y})$ กรณี $n\geq 30$ เมื่อใช้ทฤษฎีเข้าสู่ส่วนกลางได้ว่า

$$Z = \frac{\hat{Y} - \bar{Y}}{\sqrt{\hat{V}(\hat{Y})}} \sim \text{app}N(0,1) \quad (2-1)$$

เนื่องจากตัวสถิติ Z มีการแจกแจงแบบปกติโดยประมาณ และกำหนดระดับความเชื่อมั่น (ความน่าจะเป็นที่ $\bar{Y}$ จะอยู่ในช่วงความเชื่อมั่นที่สร้างขึ้น) ตามภาพที่ 2.1



ภาพที่ 2.1 ระดับความเชื่อมั่นของการประมาณค่า $\bar{Y}$

กัลยา วานิชย์บัญชา (2547) และ Casella และ Berger (2002) ได้อธิบายขั้นตอนการสร้างช่วงความเชื่อมั่นโดยอาศัยข้อมูลตามภาพที่ 2.1 ดังนี้

$$\begin{aligned}
P(-z_{1-\frac{\alpha}{2}} < Z < z_{1-\frac{\alpha}{2}}) &= 1-\alpha \\
P(-z_{1-\frac{\alpha}{2}} < \frac{\hat{\bar{Y}} - \bar{Y}}{\sqrt{\hat{V}(\hat{\bar{Y}})}} < z_{1-\frac{\alpha}{2}}) &= 1-\alpha \\
P(-z_{1-\frac{\alpha}{2}} \sqrt{\hat{V}(\hat{\bar{Y}})} < \hat{\bar{Y}} - \bar{Y} < z_{1-\frac{\alpha}{2}} \sqrt{\hat{V}(\hat{\bar{Y}})}) &= 1-\alpha \\
P\left(\hat{\bar{Y}} - z_{1-\frac{\alpha}{2}} \sqrt{\hat{V}(\hat{\bar{Y}})} < \bar{Y} < \hat{\bar{Y}} + z_{1-\frac{\alpha}{2}} \sqrt{\hat{V}(\hat{\bar{Y}})}\right) &= 1-\alpha
\end{aligned}$$

ดังนั้นช่วงความเชื่อมั่น $(1-\alpha)100\%$ ของ $\bar{Y}$ มีค่าเป็น

$$C.I._{100}(\bar{Y}) = \left(\hat{\bar{Y}} - z_{1-\frac{\alpha}{2}} \sqrt{\hat{V}(\hat{\bar{Y}})}, \hat{\bar{Y}} + z_{1-\frac{\alpha}{2}} \sqrt{\hat{V}(\hat{\bar{Y}})} \right) \quad (2-2)$$

จากสมการ (2-2) จะเห็นว่า ช่วงความเชื่อมั่นของ $\bar{Y}$ อยู่ในรูปฟังก์ชันของ $\hat{\bar{Y}}$ และ $\hat{V}(\hat{\bar{Y}})$ ซึ่งค่าดังกล่าวจะมีค่าแตกต่างกันขึ้นอยู่กับแผนการสุ่มตัวอย่างที่ใช้สุ่มตัวอย่าง s ขนาด n ต่อไปผู้วิจัยจะได้กล่าวถึงช่วงความเชื่อมั่นสำหรับประมาณค่าเฉลี่ยประชากรภายใต้การสุ่มตัวอย่างแบบง่ายและแบบใช้ความน่าจะเป็นในการเลือกไม่เท่ากันดังนี้ สมมติว่าตัวอย่าง $s = \{j_1, j_2, \dots, j_n\}$ ขนาด n เมื่อ $j_l \neq j_k$ โดยที่ $j_l, j_k \in U$ และ $l, k = 1, 2, \dots, n$ ถูกสุ่มโดยใช้แผนการสุ่มตัวอย่างแบบง่ายและไม่ใส่คืน (Simple random sampling without replacement) ตัวประมาณค่าเฉลี่ยประชากรแบบจุดเขียนแทนด้วย $\hat{\bar{Y}} = \bar{y}$ และกำหนดค่าตามที่แสดงในสมการที่ (2-3)

$$\hat{\bar{Y}} = \bar{y} = n^{-1} \sum_{i \in s} y_i \quad (2-3)$$

ภายใต้การสุ่มดังกล่าว ค่าความแปรปรวนของ $\bar{y}$ เขียนแทนด้วย $V(\bar{y})$ และกำหนดค่าตามที่แสดงในสมการที่ (2-4)

$$V(\bar{y}) = (1-f) \frac{S^2}{n} \quad (2-4)$$

เมื่อ $f = nN^{-1}$ แทนสัดส่วนตัวอย่าง (Sampling fraction) และ $S^2 = (N-1)^{-1} \sum_{i \in U} (y_i - \bar{Y})^2$ โดยปกติค่า $V(\bar{y})$ จะไม่ทราบค่า เนื่องจากค่า S^2 ขึ้นอยู่กับข้อมูลของประชากรทั้งหมด ทำให้ไม่สามารถคำนวณค่าได้ ดังนั้นจำเป็นต้องประมาณค่า $V(\bar{y})$ จากข้อมูลของตัวอย่าง ตัวประมาณค่าของ $V(\bar{y})$ เขียนแทนด้วยสัญลักษณ์ $\hat{V}(\bar{y})$ และกำหนดค่าตามที่แสดงในสมการที่ (2-5)

$$\hat{V}(\bar{y}) = (1-f) \frac{\hat{S}^2}{n} \quad (2-5)$$

เมื่อ $\hat{S}^2 = (n-1)^{-1} \sum_{i \in S} (y_i - \hat{Y})^2$ นอกจากนี้นักรณินที่ ขนาดตัวอย่างมีค่าไม่เกิน 5% ของขนาดประชากร (สัดส่วนตัวอย่างมีขนาดเล็ก) ค่า $f = nN^{-1}$ จะมีค่าเข้าใกล้ 0 ดังนั้นค่า $\hat{V}(\bar{y})$ ในสมการ (2-5) ลดรูปเป็น

$$\hat{V}(\bar{y}) = \frac{\hat{S}^2}{n} \quad (2-6)$$

จากนั้นอาศัยขั้นตอนตามที่ได้กล่าวถึงในสมการที่ (2-1) ถึง (2-2) สามารถสรุปได้ว่าภายใต้การ สุ่มตัวอย่างแบบง่ายและไม่ใส่คืน ช่วงความเชื่อที่ระดับ $(1-\alpha)100\%$ ของ $\bar{Y}$ มีค่าตามที่แสดงในสมการที่ (2-7)

$$C.I._{01}(\bar{Y}) = \left(\bar{y} - z_{1-\frac{\alpha}{2}} \frac{\hat{S}}{\sqrt{n}}, \bar{y} + z_{1-\frac{\alpha}{2}} \frac{\hat{S}}{\sqrt{n}} \right) \quad (2-7)$$

ตัวอย่าง 2.1 จากการสุ่มตัวอย่างแบบง่ายและไม่ใส่คืน ขนาด 10 ครั้วเรือน จากประชากรขนาด 100 ครั้วเรือน จากนั้นสอบถามจำนวนคนในแต่ละครั้วเรือน ได้ดังนี้

i	2	15	25	30	34	55	60	76	89	93
y_i	4	3	2	4	6	5	2	3	5	7

(สุชาดา กิระนันท์, 2542)

เมื่อ y_i แทนจำนวนคนในครั้วเรือนที่ i จากข้อมูลข้างต้นจงประมาณจำนวนคนเฉลี่ยในแต่ละครั้วเรือนที่ระดับความเชื่อมั่น 95%

วิธีทำ ที่ระดับความเชื่อมั่น 95% ได้ค่า $\alpha = 0.05$ ดังนั้นค่า $z_{1-\frac{\alpha}{2}} = z_{0.975} = 1.960$

$$\sum_{i \in S} y_i = 4 + 3 + \dots + 7 = 41$$

$$\bar{y} = n^{-1} \sum_{i \in S} y_i = 4.1$$

$$\hat{S}^2 = \frac{1}{n-1} \sum_{i \in S} (y_i - \bar{y})^2 = \frac{1}{9} [(4-4.1)^2 + \dots + (7-4.1)^2] = 2.77$$

$$\hat{S} = \sqrt{2.77} = 1.66$$

ช่วงความเชื่อมั่นของ $\bar{Y}$ ที่ระดับ 95% สามารถคำนวณได้จาก

$$\begin{aligned} \bar{y} \pm z_{1-\frac{\alpha}{2}} \frac{\hat{S}}{\sqrt{n}} &= 4.1 \pm 1.96 \left(\frac{1.66}{\sqrt{10}} \right) \\ &= 4.1 \pm 1.96 \left(\frac{1.66}{3.16} \right) \\ &= 4.1 \pm 1.03 \end{aligned}$$

$$=(4.1-1.03, 4.1+1.03)$$

$$=(3.07, 5.13)$$

ดังนั้น $C.I_{.01}(\bar{Y})=(3.07, 5.13)$ หรือจำนวนคนเฉลี่ยในแต่ละครัวเรือนมีค่าตั้งแต่ 3.07 คน ถึง 5.13 คน ที่ระดับความเชื่อมั่น 95%

ตัวอย่างที่ 2.1 ผู้วิจัยได้แสดงวิธีการคำนวณช่วงความเชื่อมั่นสำหรับการประมาณค่า $\bar{Y}$ ภายใต้การสุ่มตัวอย่างแบบง่ายและไม่ใส่คืน ต่อไปผู้วิจัยจะได้กล่าวถึงแผนการสุ่มตัวอย่างที่ต่างจากแผนการสุ่มแบบง่ายและไม่ใส่คืน เนื่องจากความน่าจะเป็นที่แต่ละหน่วยของประชากรจะถูกเลือกเป็นสมาชิกของตัวอย่างมีค่าไม่เท่ากัน และเรียกแผนการสุ่มตัวอย่างดังกล่าวว่า แผนการสุ่มตัวอย่างแบบใช้ความน่าจะเป็นในการเลือกไม่เท่ากัน ตัวอย่างเช่น แผนการสุ่มตัวอย่างของ Midzuno (1952) แผนการสุ่มตัวอย่างแบบ Poisson เป็นต้น อย่างไรก็ตามแผนการสุ่มดังกล่าวจำเป็นต้องทราบข้อมูลของตัวแปรช่วยที่มีความสัมพันธ์กับตัวแปรที่สนใจศึกษา ก่อนเริ่มทำการสุ่มตัวอย่าง กำหนดให้ k แทนตัวแปรช่วย และ k_i แทนค่าสังเกตหน่วยที่ i ของ k เมื่อ $i \in U$ กรณีที่เป็นการสุ่มแบบใส่คืน Hansen และ Hurwitz (1946) ได้นำเสนอตัวประมาณค่าเฉลี่ยประชากร เขียนแทนด้วยสัญลักษณ์ $\hat{Y}_{HH}$ กำหนดค่าตามที่แสดงในสมการ (2-8)

$$\hat{Y}_{HH} = \frac{1}{nN} \sum_{i \in S} \frac{y_i}{P_i} = \frac{1}{N} \sum_{i \in S} f(y_i) = \frac{1}{N} \sum_{i \in S} b_i \quad (2-8)$$

เมื่อ $b_i = f(y_i) = \frac{1}{n} \frac{y_i}{P_i}$ ค่า $P_i = \frac{k_i}{K}$ และ $K = \sum_{i \in U} k_i$ ส่วนตัวประมาณค่าความแปรปรวนของ $\hat{Y}_{HH}$ หรือ $\hat{V}(\hat{Y}_{HH})$ มีค่าตามที่แสดงในสมการที่ (2-9)

$$\hat{V}(\hat{Y}_{HH}) = \frac{1}{N^2} \left(\frac{n}{n-1} \right) \left[\sum_{i \in S} \hat{b}_i^2 - \frac{1}{n} \left(\sum_{i \in S} \hat{b}_i \right)^2 \right] \quad (2-9)$$

เมื่อ $\hat{b}_i = b_i = \frac{1}{n} \frac{y_i}{P_i}$ จากนั้นอาศัยขั้นตอนตามที่ได้กล่าวถึงในสมการที่ (2-1) ถึง (2-2) สามารถสรุปได้ว่าภายใต้การสุ่มตัวอย่างแบบใช้ความน่าจะเป็นในการเลือกไม่เท่ากันและใส่คืน ช่วงความเชื่อที่ระดับ $(1-\alpha)100\%$ ของ $\bar{Y}$ มีค่าตามที่แสดงในสมการที่ (2-10)

$$C.I_{.02}(\bar{Y}) = \left(\hat{Y}_{HH} - z_{1-\frac{\alpha}{2}} \sqrt{\hat{V}(\hat{Y}_{HH})}, \hat{Y}_{HH} + z_{1-\frac{\alpha}{2}} \sqrt{\hat{V}(\hat{Y}_{HH})} \right) \quad (2-10)$$

เมื่อ $\hat{Y}_{HH}$ และ $\hat{V}(\hat{Y}_{HH})$ กำหนดค่าตามสมการ (2-8) และ (2-9) ตามลำดับ และค่า $z_{1-\frac{\alpha}{2}}$ กำหนดค่าตามสมการ (1-5)

ตัวอย่าง 2.2 กำหนดข้อมูลประชากรตามการศึกษาของ Lohr (1999) ชื่อไฟล์ “*agstrat.dat*” ประกอบด้วยประชากรขนาด $N=298$ ตัวแปรที่สนใจศึกษาได้แก่ $y = ACRES92$ แทนจำนวนพื้นที่ที่ใช้ในการทำฟาร์มในปี 1992 ตัวแปรช่วยได้แก่ $k = ACRES82$ แทนจำนวนพื้นที่ที่ใช้ในการทำฟาร์มในปี 1982 จากข้อมูลประชากรดังกล่าวสุ่มตัวอย่างขนาด $n=30$ โดยใช้แผนการสุ่มตัวอย่างแบบใช้ความน่าจะเป็นในการเลือกไม่เท่ากันและใส่คืน ได้ข้อมูลตัวอย่างดังนี้

ตัวอย่างที่	i	y_i	k_i	P_i
1	2	124,694	139,111	0.0015
2	17	169,622	179,044	0.002
3	25	403,375	382,477	0.0042
4	47	195,077	196,321	0.0021
5	54	224,030	240,591	0.0026
6	56	16,099	17,700	0.0002
7	65	241,778	274,368	0.003
8	82	249,731	260,395	0.0028
9	88	312,079	331,520	0.0036
10	97	451,362	464,739	0.0051
11	100	182,572	193,237	0.0021
12	101	592,207	602,726	0.0066
13	118	192,116	297,071	0.0032
14	126	346,653	350,350	0.0038
15	127	545,666	507,523	0.0055
16	131	24,239	33,762	0.0004
17	132	21,975	42,527	0.0005
18	133	138,573	177,061	0.0019
19	137	151,743	188,333	0.0021
20	158	662,121	659,095	0.0072
21	169	38,313	41,585	0.0005
22	186	593,819	612,836	0.0067
23	201	561,521	513,862	0.0056
24	208	83,218	73,766	0.0008
25	241	9,910	13,613	0.0001
26	246	219,444	235,409	0.0026
27	254	295,601	256,837	0.0028

ตัวอย่างที่	i	y_i	k_i	P_i
28	263	399,011	315,061	0.0034
29	264	1,334,041	1,353,774	0.0148
30	282	213,004	201,657	0.0022

จากข้อมูลข้างต้นหาช่วงความเชื่อมั่นของจำนวนพื้นที่เฉลี่ยที่ใช้ในการทำฟาร์มในปี 1992 ที่ระดับความเชื่อมั่น 90%

วิธีทำ ที่ระดับความเชื่อมั่น 90% ได้ค่า $\alpha=0.10$ ดังนั้นค่า $z_{1-\frac{\alpha}{2}} = z_{0.950} = 1.645$

จากข้อมูลข้างต้นสามารถคำนวณค่าสถิติที่ใช้ในการคำนวณช่วงความเชื่อมั่นได้ดังนี้

$$\begin{aligned}\hat{Y}_{HH} &= \frac{1}{nN} \sum_{i \in S} \frac{y_i}{P_i} \\ &= \frac{1}{(30)(298)} \left[\frac{124,694}{0.0015} + \dots + \frac{213,004}{0.0022} \right] \\ &= \frac{1}{8,940} (2,600,287,266) = 290,859.87\end{aligned}$$

$$\begin{aligned}\hat{V}(\hat{Y}_{HH}) &= \frac{1}{N^2} \left(\frac{n}{n-1} \right) \left[\sum_{i \in S} \hat{b}_i^2 - \frac{1}{n} \left(\sum_{i \in S} \hat{b}_i \right)^2 \right] \\ &= \frac{1}{N^2} \left(\frac{n}{n-1} \right) \left[\sum_{i \in S} \left(\frac{1}{n} \frac{y_i}{P_i} \right)^2 - \frac{1}{n} \left(\sum_{i \in S} \frac{1}{n} \frac{y_i}{P_i} \right)^2 \right] \\ &= \frac{1}{N^2} \frac{1}{n(n-1)} \left[\sum_{i \in S} \left(\frac{y_i}{P_i} \right)^2 - \frac{1}{n} \left(\sum_{i \in S} \frac{y_i}{P_i} \right)^2 \right]\end{aligned}$$

ดังนั้น

$$\hat{V}(\hat{Y}_{HH}) = \frac{1}{N^2} \frac{1}{n(n-1)} \left[\sum_{i \in S} \left(\frac{y_i}{P_i} \right)^2 - \frac{1}{n} \left(\sum_{i \in S} \frac{y_i}{P_i} \right)^2 \right] \quad (2-11)$$

จากข้อมูลของตัวอย่างตามที่แสดงในตารางข้างต้น สามารถคำนวณค่า

$$\sum_{i \in S} \frac{y_i^2}{P_i^2} = 231,601,152,064,328,000$$

$$\sum_{i \in S} \frac{y_i}{P_i} = 2,600,287,266$$

แทนค่า $\sum_{i \in S} \frac{y_i}{P_i}$ และ $\sum_{i \in S} \frac{y_i^2}{P_i^2}$ ในสมการที่ (2-11) ได้ว่า

$$\hat{V}(\hat{Y}_{HH}) = \frac{1}{N^2} \frac{1}{n(n-1)} \left[\sum_{i \in S} \left(\frac{y_i}{P_i} \right)^2 - \frac{1}{n} \left(\sum_{i \in S} \frac{y_i}{P_i} \right)^2 \right]$$

$$= \frac{1}{(298)^2} \frac{1}{30(29)} \left(231,601,152,064,328,000 - \frac{1}{30} (2,600,287,266)^2 \right)$$

$$= 80,482,332.32$$

ช่วงความเชื่อมั่นของ $\bar{Y}$ ที่ระดับ 90% สามารถคำนวณได้จาก

$$\hat{Y}_{HH} \pm z_{1-\frac{\alpha}{2}} \sqrt{\hat{V}(\hat{Y}_{HH})} = 290,859.87 \pm 1.645 \sqrt{80,482,332.32}$$

$$= 290,859.87 \pm 1.645(8,971.19)$$

$$= 290,859.87 \pm 14,757.62$$

$$= (290,859.87 - 14,757.62, 290,859.87 + 14,757.62)$$

$$= (276,102.25, 305,617.49)$$

ดังนั้น $C.I._{0.2}(\bar{Y}) = (276,102.25, 305,617.49)$ หรือ จำนวนพื้นที่เฉลี่ยที่ใช้ในการทำฟาร์มในปี 1992 ที่ระดับความเชื่อมั่น 90% มีค่าตั้งแต่ 276,102.25 เอเคอร์ ถึง 305,617.49 เอเคอร์

Hansen และ Hurwitz (1946) ได้นำเสนอตัวประมาณค่า และช่วงความเชื่อมั่นสำหรับประมาณค่าเฉลี่ยประชากร ภายใต้แผนการสุ่มตัวอย่างแบบใช้ความน่าจะเป็นในการเลือกไม่เท่ากัน และใส่คืน ซึ่งแผนการสุ่มตัวอย่างดังกล่าวมีโอกาสที่หน่วยตัวอย่างเดียวกัน จะถูกสุ่มเลือกมากกว่า 1 ครั้ง ซึ่งอาจส่งผลต่อประสิทธิภาพในการประมาณค่า ดังนั้น Horvitz และ Thompson (1952) ได้พัฒนาตัวประมาณค่าที่ Hansen และ Hurwitz (1946) นำเสนอไปยังแผนการสุ่มแบบใช้ความน่าจะเป็นในการเลือกไม่เท่ากันและไม่ใส่คืน ซึ่งตัวประมาณค่าที่ Horvitz และ Thompson (1952) นำเสนออยู่ในรูปตัวประมาณค่าแบบเชิงเส้น เขียนแทนด้วย $\hat{Y}_{HT}$ และมีค่าตามที่แสดงในสมการที่ (2-14)

$$\hat{Y}_{HT} = \frac{1}{N} \sum_{i \in s} \frac{y_i}{\pi_i} \quad (2-14)$$

เมื่อ $\pi_i = P(i \in s)$ และตัวประมาณค่าความแปรปรวนของ $\hat{Y}_{HT}$ มีค่าตามที่แสดงในสมการที่ (2-15)

$$\hat{V}(\hat{Y}_{HT}) = \frac{1}{N^2} \left[\sum_{i \in s} \frac{1 - \pi_i}{\pi_i^2} y_i^2 + \sum_{i \in s} \sum_{j \in s, j \neq i} \frac{\pi_{ij} - \pi_i \pi_j}{\pi_{ij} \pi_i \pi_j} y_i y_j \right] \quad (2-15)$$

เมื่อ $\pi_{ij} = P(i \wedge j \in s)$ นอกจากนี้เมื่ออาศัยวิธีการตามที่ได้กล่าวถึงในสมการที่ (2-1) และ (2-2) ช่วงความเชื่อมั่น $(1-\alpha)100\%$ ของ $\bar{Y}$ ที่สร้างจากตัวประมาณค่าของ Horvitz และ Thompson (1952) มีค่าตามที่แสดงในสมการที่ (2-16)

$$C.I._{(1)}(\bar{Y}) = \left(\hat{Y}_{HT} - z_{1-\frac{\alpha}{2}} \sqrt{\hat{V}(\hat{Y}_{HT})}, \hat{Y}_{HT} + z_{1-\frac{\alpha}{2}} \sqrt{\hat{V}(\hat{Y}_{HT})} \right) \quad (2-16)$$

นอกจากตัวประมาณค่าแบบเชิงเส้นของ Horvitz และ Thompson (1952) ในปี 1971 Hájek ได้นำเสนอตัวประมาณค่าแบบไม่เป็นเชิงเส้นเพื่อใช้ประมาณค่าเฉลี่ยประชากร เขียนแทนด้วย สัญลักษณ์ $\hat{Y}_{Hajek}$ และมีค่าตามที่แสดงในสมการ (2-17)

$$\hat{Y}_{Hajek} = \frac{\sum_{i \in S} \frac{y_i}{\pi_i}}{\sum_{i \in S} \frac{1}{\pi_i}} \quad (2-17)$$

เมื่อ π_i กำหนดค่าตามสมการ (2-14) และ π_{ij} กำหนดค่าตามสมการ (2-15) สำหรับคุณสมบัติของ $\hat{Y}_{HT}$ และ $\hat{Y}_{Hajek}$ Särndal , Swenson และ Wretman (1992) กล่าวว่าในกรณีที่ตัวแปรที่สนใจศึกษา y_i มีความสัมพันธ์ในทิศทางตรงข้ามกับ π_i ตัวประมาณค่า $\hat{Y}_{Hajek}$ จะมีประสิทธิภาพดีกว่า $\hat{Y}_{HT}$ ต่อมา Tillé และ Matei (2016) ได้สนับสนุนคำกล่าวของ Särndal , Swenson และ Wretman (1992) โดยการใช้การจำลองข้อมูลด้วยวิธีมอนติคาร์โล และเปรียบเทียบประสิทธิภาพของ $\hat{Y}_{HT}$ และ $\hat{Y}_{Hajek}$ จากค่าเฉลี่ยของความคลาดเคลื่อนยกกำลังสอง (Mean Square Error) อย่างไรก็ตาม ตัวประมาณค่าของ Hájek (1971) หรือ $\hat{Y}_{Hajek}$ จะอยู่ในรูปตัวประมาณค่าแบบไม่เป็นเชิงเส้น ดังนั้นการหาค่าคลาดหมาย หรือค่าความแปรปรวนจำเป็นต้องใช้วิธีเทเลอร์แบบเชิงเส้นเพื่อปรับตัวประมาณค่าดังกล่าวให้อยู่ในรูปแบบเชิงเส้น จากนั้นประมาณค่าความแปรปรวนของตัวประมาณค่าที่ไม่เป็นเชิงเส้น จากตัวประมาณค่าแบบเชิงเส้น หลังจากที่ใช้วิธีดังกล่าว ตัวประมาณค่าความแปรปรวนของ $\hat{Y}_{Hajek}$ มีค่าตามที่แสดงในสมการ (2-18)

$$\hat{V}(\hat{Y}_{Ha'jek}) = \frac{1}{N^2} \left[\sum_{i \in S} \frac{1-\pi_i}{\pi_i^2} (y_i - \hat{Y}_{Ha'jek})^2 + \sum_{i \in S} \sum_{j \in S, j \neq i} \frac{\pi_{ij} - \pi_i \pi_j}{\pi_{ij} \pi_i \pi_j} (y_i - \hat{Y}_{Ha'jek})(y_j - \hat{Y}_{Ha'jek}) \right] \quad (2-18)$$

เมื่อ $\hat{Y}_{Ha'jek}$ กำหนดค่าตามที่แสดงในสมการ (2-17) จากนั้นอาศัยวิธีการตามที่ได้กล่าวถึงในสมการที่ (2-1) และ (2-2) ช่วงความเชื่อมั่น $(1-\alpha)100\%$ ของ $\bar{Y}$ ที่สร้างจากตัวประมาณค่า Hájek (1971) มีค่าตามที่แสดงในสมการที่ (2-19)

$$C.I_{(2)}(\bar{Y}) = \left(\hat{Y}_{Ha'jek} - z_{1-\frac{\alpha}{2}} \sqrt{\hat{V}(\hat{Y}_{Ha'jek})}, \hat{Y}_{Ha'jek} + z_{1-\frac{\alpha}{2}} \sqrt{\hat{V}(\hat{Y}_{Ha'jek})} \right) \quad (2-19)$$

ตัวอย่าง 2.3 จากตัวอย่าง 2.2 สมมติว่าสุ่มตัวอย่าง s ขนาด $n=3$ โดยใช้แผนการสุ่มตัวอย่างของ Midzuno (1952) และได้หน่วยตัวอย่างตามที่แสดงในตารางข้างล่าง

ลำดับที่	1	2	3
หน่วยตัวอย่างที่	2	65	282
y_i	124,694	241,778	213,004
P_i	0.0015	0.0030	0.0022

จากข้อมูลข้างต้นจึงประมาณจำนวนพื้นที่ที่ใช้ในการทำฟาร์มของเกษตรกร ในปี 1992 ด้วยช่วงความเชื่อมั่นที่สร้างจากตัวประมาณค่าของ Horvitz และ Thompson (1952) ที่ระดับความเชื่อมั่น 99%

วิธีทำ ที่ระดับความเชื่อมั่น 99% ได้ค่า $\alpha=0.01$ ดังนั้นค่า

$$z_{1-\frac{\alpha}{2}} = z_{0.995} = 2.576 \quad (2-20)$$

ส่วนค่า $N=298$ และ $n=3$ ภายใต้แผนการสุ่มตัวอย่างของ Midzuno (1952) ค่า π_i และ π_{ij} กำหนดค่าตามที่แสดงในสมการที่ (2-20) และ (2-21) ตามลำดับ

$$\pi_i = P_i \frac{N-n}{N-1} + \frac{n-1}{N-1} \quad (2-21)$$

$$\pi_{ij} = (P_i + P_j) \frac{N-n}{N-1} \frac{n-1}{N-2} + \frac{n-1}{N-1} \frac{n-2}{N-2} \quad (2-22)$$

ค่า π_i เมื่อ $i \in s = \{2, 65, 282\}$ คำนวณค่าได้ดังนี้ จากสมการ (2-21)

$$\pi_i = P_i \frac{N-n}{N-1} + \frac{n-1}{N-1} = P_i \frac{298-3}{298-1} + \frac{3-1}{298-1} = P_i \frac{295}{297} + \frac{2}{297}$$

กรณีที่ $i=2$

$$\pi_2 = P_2 \frac{295}{297} + \frac{2}{297} = (0.0015) \frac{295}{297} + \frac{2}{297} = 0.0082239$$

กรณีที่ $i=65$

$$\pi_{65} = P_{65} \frac{295}{297} + \frac{2}{297} = (0.003) \frac{295}{297} + \frac{2}{297} = 0.0097138$$

กรณีที่ $i=282$

$$\pi_{282} = P_{282} \frac{295}{297} + \frac{2}{297} = (0.0022) \frac{295}{297} + \frac{2}{297} = 0.0089192$$

สำหรับค่า π_{ij} สามารถคำนวณค่าได้ดังนี้

$$\begin{aligned}\pi_{ij} = \pi_{ji} &= (P_i + P_j) \frac{298-3}{298-1} \frac{3-1}{298-2} + \frac{3-1}{298-1} \frac{3-2}{298-2} \\ &= (P_i + P_j) \frac{295}{297} \frac{2}{296} + \frac{2}{297} \frac{1}{296}\end{aligned}$$

กรณีนี้ที่ $i=2$ และ $j=65$ หรือ $i=65$ และ $j=2$

$$\begin{aligned}\pi_{ij} &= (P_2 + P_{65}) \frac{298-3}{298-1} \frac{3-1}{298-2} + \frac{3-1}{298-1} \frac{3-2}{298-2} \\ &= (0.0015 + 0.0030) \frac{295}{297} \frac{2}{296} + \frac{2}{297} \frac{1}{296} = 0.0000530\end{aligned}$$

กรณีนี้ที่ $i=2$ และ $j=282$ หรือ $i=282$ และ $j=2$

$$\begin{aligned}\pi_{ij} &= (P_2 + P_{282}) \frac{295}{297} \frac{2}{296} + \frac{2}{297} \frac{1}{296} \\ &= (0.0015 + 0.0022) \frac{295}{297} \frac{2}{296} + \frac{2}{297} \frac{1}{296} = 0.0000476\end{aligned}$$

กรณีนี้ที่ $i=65$ และ $j=282$ หรือ $i=282$ และ $j=65$

$$\begin{aligned}\pi_{ij} &= (P_{65} + P_{282}) \frac{295}{297} \frac{2}{296} + \frac{2}{297} \frac{1}{296} \\ &= (0.0030 + 0.0022) \frac{295}{297} \frac{2}{296} + \frac{2}{297} \frac{1}{296} = 0.0000576\end{aligned}$$

ดังนั้นค่า $\hat{Y}_{HT}$ มีค่าดังนี้

$$\begin{aligned}\hat{Y}_{HT} &= \frac{1}{N} \sum_{i \in S} \frac{y_i}{\pi_i} \\ &= \frac{1}{298} \left[\frac{124,694}{0.0082239} + \frac{241,778}{0.0097138} + \frac{213,004}{0.0089192} \right] \\ &= 214,543.84\end{aligned}\tag{2-23}$$

จากสมการ (2-15) ค่า $\hat{V}(\hat{Y}_{HT})$ มีค่าดังนี้

$$\hat{V}(\hat{Y}_{HT}) = \frac{1}{N^2} \left[\sum_{i \in S} \frac{1-\pi_i}{\pi_i^2} y_i^2 + \sum_{i \in S} \sum_{j \in S, j \neq i} \frac{\pi_{ij} - \pi_i \pi_j}{\pi_{ij} \pi_i \pi_j} y_i y_j \right]\tag{2-24}$$

จากสมการ (2-24)

$$\begin{aligned}\sum_{i \in S} \frac{1-\pi_i}{\pi_i^2} y_i^2 &= \frac{1-\pi_2}{\pi_2^2} y_2^2 + \frac{1-\pi_{65}}{\pi_{65}^2} y_{65}^2 + \frac{1-\pi_{282}}{\pi_{282}^2} y_{282}^2 \\ &= 1,406,749,329,883,040\end{aligned}\tag{2-25}$$

$$\sum_{i \in S} \sum_{j \in S, j \neq i} \frac{\pi_{ij} - \pi_i \pi_j}{\pi_{ij} \pi_i \pi_j} y_i y_j = -1,374,013,996,175,610.00\tag{2-26}$$

แทนค่า (2-25) และ (2-26) ในสมการที่ (2-24) ได้ว่า

$$\begin{aligned}\hat{V}(\hat{Y}_{HT}) &= \frac{1}{298^2} (1,406,749,329,883,040 - 1,374,013,996,175,610.00) \\ &= 368,624,540.64\end{aligned}\quad (2-27)$$

จากสมการ (2-16) ช่วงความเชื่อมั่นที่สร้างจากตัวประมาณค่าของ Horvitz และ Thompson (1952) มีค่าดังนี้

$$C.I._{(1)}(\bar{Y}) = \left(\hat{Y}_{HT} - z_{1-\frac{\alpha}{2}} \sqrt{\hat{V}(\hat{Y}_{HT})}, \hat{Y}_{HT} + z_{1-\frac{\alpha}{2}} \sqrt{\hat{V}(\hat{Y}_{HT})} \right) \quad (2-28)$$

แทนค่า (2-20) (2-23) และ (2-27) ใน (2-28) ได้ว่า

$$C.I._{(1)}(\bar{Y}) = (165,085.67, 264,002.00)$$

ดังนั้น $C.I._{(1)}(\bar{Y}) = (165,085.67, 264,002.00)$ หรือ จำนวนพื้นที่เฉลี่ยที่ใช้ในการทำฟาร์มในปี 1992 ที่ระดับความเชื่อมั่น 99% มีค่าตั้งแต่ 165,085.67 เอเคอร์ ถึง 264,002.00 เอเคอร์

2.2 เกณฑ์ที่ใช้ในการเปรียบเทียบประสิทธิภาพช่วงความเชื่อมั่น

หัวข้อนี้ผู้วิจัยจะได้กล่าวถึง ค่ากว้างเฉลี่ยของช่วงความเชื่อมั่น เพื่อใช้เป็นเกณฑ์ในการเปรียบเทียบประสิทธิภาพช่วงความเชื่อมั่น ซึ่งค่าดังกล่าวสามารถคำนวณได้จากการนำผลรวมความกว้างของช่วงความเชื่อมั่นหารด้วยจำนวนรอบของการทำซ้ำและมีค่าตามที่แสดงในสมการ (2-29)

$$AL = \frac{1}{M} \sum_{i=1}^M (U_i - L_i) \quad (2-29)$$

เมื่อ L_i แทนขีดความเชื่อมั่นล่าง ของช่วงความเชื่อมั่น

U_i แทนขีดความเชื่อมั่นบน ของช่วงความเชื่อมั่น

M แทน จำนวนรอบของการทำซ้ำ

ตัวอย่าง 2.4 จากตัวอย่าง 2.2 สมมติว่าสุ่มตัวอย่าง s ขนาด $n=3$ โดยใช้แผนการสุ่มตัวอย่างของ Midzuno (1952) กำหนดการทำซ้ำทั้งหมด 4 รอบ และได้ข้อมูลตัวอย่างดังนี้

รอบที่	1	2	3	4
s	{6,65,282}	{10,178,220}	{68,110,280}	{5,190,200}

จากข้อมูลข้างต้นคำนวณหา ค่ากว้างเฉลี่ยของช่วงความเชื่อมั่นของจำนวนพื้นที่ที่ใช้ในการทำฟาร์มของเกษตรกรที่สร้างจากตัวประมาณของ Horvitz และ Thompson (1952) ที่ระดับความเชื่อมั่น 99%

วิธีทำ โดยอาศัยขั้นตอนตามที่แสดงในตัวอย่าง 2.3 สามารถคำนวณค่าสถิติที่เกี่ยวข้องได้ดังนี้

รอบที่	ตัวอย่าง	L_i	U_i	$U_i - L_i$
1	$s = \{6, 65, 282\}$	165,085.67	264,002.00	98,916.33
2	$s = \{10, 178, 220\}$	143,183.67	215,799.28	72,615.61
3	$s = \{68, 110, 280\}$	44,715.16	851,181.04	806,465.88
4	$s = \{5, 190, 200\}$	9,663.09	538,561.73	528,898.64
รวม				$\sum_{i=1}^4 (U_i - L_i) = 1,506,896.46$

ค่ากว้างเฉลี่ยของช่วงความเชื่อมั่น คำนวณได้จาก

$$AL = \frac{1}{M} \sum_{i=1}^M (U_i - L_i) = \frac{1}{4} \sum_{i=1}^4 (U_i - L_i) = \frac{1}{4} (1,506,896.46) = 376,724.12$$

ดังนั้นค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่น มีค่าเท่ากับ 376,724.12

2.3 การหาความแปรปรวนของตัวประมาณค่าแบบไม่เป็นเชิงเส้น

หัวข้อ 2.1.2 ผู้วิจัยได้กล่าวถึงตัวประมาณค่าเฉลี่ยประชากร ของ Horvitz และ Thompson (1952) และตัวประมาณค่าของ Hajek (1971) ซึ่งตัวประมาณค่าที่ Horvitz และ Thompson (1952) นำเสนออยู่ในรูปแบบเชิงเส้นดังนั้นการหาค่าความแปรปรวน สามารถหาได้โดยตรงตามขั้นตอนที่แสดงใน สมการ (2-14) และ (2-15) ส่วนตัวประมาณค่าของ Hajek (1971) อยู่ในรูปของอัตราส่วน ดังนั้นตัวประมาณค่าของ Hajek (1971) เป็นตัวประมาณค่าแบบไม่เป็นเชิงเส้น การหาค่าความแปรปรวนของตัวประมาณค่าแบบที่อยู่ในลักษณะดังกล่าวมีหลายวิธีเช่น วิธีบูตสแตรป (Bootstrap Method) วิธีแจ๊คไนฟ์ (Jackknife Method) หรือวิธีเทเลอร์แบบเชิงเส้น (Taylor linearization method) สำหรับงานวิจัยในครั้งนี้ผู้วิจัยใช้ วิธีเทเลอร์แบบเชิงเส้นซึ่งมีขั้นตอนดังนี้

(1) ปรับตัวประมาณค่าแบบไม่เป็นเชิงเส้น ให้เป็นตัวประมาณค่าแบบเชิงเส้น โดยใช้วิธีเทเลอร์แบบเชิงเส้น

กำหนดให้ $\hat{Y}_{NL}$ เป็นตัวประมาณค่าแบบไม่เป็นเชิงเส้น และสามารถเขียนให้อยู่ในรูปของฟังก์ชันปรับเรียบได้ตามที่แสดงในสมการที่ (2-30)

$$\hat{Y}_{NL} = \psi(\hat{t}) \quad (2-30)$$

เมื่อ ψ แทนฟังก์ชันปรับเรียบ $\hat{t} = [\hat{t}_1 \ \hat{t}_2 \ \hat{t}_3 \ \dots \ \hat{t}_k]'$ แทนเวกเตอร์ของตัวแปรสุ่มขนาด k และ $\hat{t}_i$ เมื่อ $i=1,2,\dots,k$ แทนฟังก์ชันของตัวแปรสุ่มที่อยู่ในรูปของตัวประมาณค่าแบบ Horvitz และ Thompson (1952) เช่น $\hat{t}_1 = \sum_{i \in S} y_i / \pi_i$ และ $\hat{t}_2 = \sum_{i \in S} 1 / \pi_i$ กำหนดให้ $t = E(\hat{t}) = [t_1 \ t_2 \ t_3 \ \dots \ t_k]'$ แทนค่าคาดหวังของ $\hat{t}$ และ $t_i = E(\hat{t}_i)$ เมื่อ $i=1,2,\dots,k$ กำหนดให้ $\hat{Y}_{NL,lin}$ เป็นตัวประมาณค่าแบบเชิงเส้นของ $\hat{Y}_{NL}$ และมีค่าตามที่แสดงในสมการที่ (2-31)

$$\hat{Y}_{NL,lin} \approx \psi(t) + \sum_{i=1}^k \frac{\partial \psi(\hat{t})}{\partial \hat{t}_i} \bigg|_{\hat{t}=t} (\hat{t}_i - t_i) \quad (2-31)$$

(2) หาค่าความแปรปรวนของ $\hat{Y}_{NL}$ ตามที่แสดงในสมการที่ (2-32)

$$V(\hat{Y}_{NL}) \approx V(\hat{Y}_{NL,lin}) \quad (2-32)$$

ตัวอย่างที่ 2.5 พิจารณาตัวอย่าง s ขนาด n ถูกสุ่มตามข้อตกลงของแผนการสุ่มตัวอย่าง กำหนดให้

$$\hat{\bar{Y}}_{Hajek} = \sum_{i \in S} \frac{y_i}{\pi_i} / \sum_{i \in S} \frac{1}{\pi_i} \text{ แทนตัวประมาณค่าเฉลี่ยประชากร การหาค่าความแปรปรวนของ } \hat{\bar{Y}}_{Hajek} \text{ เป็นดังนี้}$$

(1) $\hat{\bar{Y}}_{Hajek}$ สามารถเขียนให้อยู่ในรูปของฟังก์ชันปรับเรียบ ตามที่แสดงในสมการที่ (2-33)

$$\hat{\bar{Y}}_{Hajek} = \sum_{i \in S} \frac{y_i}{\pi_i} / \sum_{i \in S} \frac{1}{\pi_i} = \frac{\hat{t}_1}{\hat{t}_2} = \psi(\hat{t}) \quad (2-33)$$

เมื่อ $\hat{t} = [\hat{t}_1 \ \hat{t}_2]'$ $= \left[\sum_{i \in S} \frac{y_i}{\pi_i} \ \sum_{i \in S} \frac{1}{\pi_i} \right]'$ ค่าคาดหวังของ $\hat{t}$ มีค่าเป็น $t = E(\hat{t}) = [t_1 \ t_2]'$ $= \left[\sum_{i \in U} y_i \ N \right]'$

เมื่อใช้วิธีเทเลอร์แบบเชิงเส้น ตัวประมาณแบบเชิงเส้นของ $\hat{\bar{Y}}_{Hajek}$ เขียนแทนด้วย $\hat{\bar{Y}}_{Hajek,lin}$ มีค่าเป็น

$$\hat{\bar{Y}}_{Hajek,lin} \approx \text{Constant} + \frac{1}{N} \sum_{i \in S} \frac{1}{\pi_i} (y_i - \bar{Y}) \quad (2-34)$$

(2) ความแปรปรวนของ $\hat{\bar{Y}}_{Hajek}$ เขียนแทนด้วย $V(\hat{\bar{Y}}_{Hajek})$ กำหนดค่าตามที่แสดงในสมการ (2-35)

$$\begin{aligned} V(\hat{\bar{Y}}_{Hajek}) &\approx V(\hat{\bar{Y}}_{Hajek,lin}) \\ &= V\left(\text{Constant} + \frac{1}{N} \sum_{i \in S} \frac{1}{\pi_i} (y_i - \bar{Y})\right) = \frac{1}{N^2} V\left(\sum_{i \in S} \frac{1}{\pi_i} (y_i - \bar{Y})\right) \\ &= \frac{1}{N^2} \left[\sum_{i \in U} \left(\frac{1 - \pi_i}{\pi_i} \right) (y_i - \bar{Y})^2 + \sum_{i \in U} \sum_{j \neq i \in U} \left(\frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j} \right) (y_i - \bar{Y})(y_j - \bar{Y}) \right] \end{aligned} \quad (2-35)$$

ตัวประมาณของ $V(\hat{\bar{Y}}_{Hajek})$ เขียนแทนด้วย $\hat{V}(\hat{\bar{Y}}_{Hajek})$ และกำหนดค่าตามที่แสดงในสมการ (2-36)

$$\hat{V}(\hat{Y}_{Hajek}) = \frac{1}{N^2} \left[\sum_{i \in S} D_i (y_i - \hat{Y}_{Hajek})^2 + \sum_{i \in S} \sum_{j \neq i \in S} D_{ij} (y_i - \hat{Y}_{Hajek})(y_j - \hat{Y}_{Hajek}) \right] \quad (2-36)$$

$$\text{เมื่อ } D_i = \frac{1 - \pi_i}{\pi_i^2} \text{ ค่า } D_{ij} = \frac{\pi_{ij} - \pi_i \pi_j}{\pi_{ij} \pi_i \pi_j} \text{ และ } \hat{Y}_{Hajek} = \sum_{i \in S} \frac{y_i}{\pi_i} / \sum_{i \in S} \frac{1}{\pi_i}$$

2.4 งานวิจัยที่เกี่ยวข้อง

กัญญ์พิชญา พุทธะไชยทัศน์ (2552) ศึกษาการเปรียบเทียบวิธีการประมาณการแจกแจงของข้อมูลสมมาตรและไม่สมมาตรด้วยวิธีแจ็กไนฟ์ กับ วิธีบูตสเตรป การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของวิธีประมาณการแจกแจงระหว่างวิธีแจ็กไนฟ์ กับ วิธีบูตสเตรป โดยจะประมาณพารามิเตอร์ค่าเฉลี่ย ความแปรปรวน ความเบ้ และความโด่ง ทั้งการประมาณแบบจุด และการประมาณแบบช่วง ข้อมูลที่ใช้ในการวิจัยได้จากการจำลองโดยเทคนิคมอนติคาร์โล สำหรับการแจกแจงปกติปลอมปน มีพารามิเตอร์กำหนดความแปรปรวน (σ^2) เท่ากับ 25 เปอร์เซนต์การปลอมปน (p) เท่ากับ 10% และ 30% และสเกลแฟคเตอร์ (c) เท่ากับ 5 และ 10 สำหรับการแจกแจงชีก้ามีค่าพารามิเตอร์ β เท่ากับ 0.1, 0.5, 1 และ 1.5 สำหรับการแจกแจงแกมมา มีค่าพารามิเตอร์ α เท่ากับ 3, 4, 6 และ 8 ค่าพารามิเตอร์ β เท่ากับ 0.1 โดยกำหนดขนาดตัวอย่าง เท่ากับ 100, 200, 300, 400, 500, 600, 700, 800, 900 และ 1,000 การจำลองข้อมูลกระทำซ้ำ 500 ครั้ง ในแต่ละสถานการณ์ ค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง และความเอนเอียง เป็นเกณฑ์ที่ใช้ในการเปรียบเทียบการประมาณค่าแบบจุด และสัมประสิทธิ์ความเชื่อมั่น เป็นเกณฑ์ที่ใช้ในการเปรียบเทียบการประมาณค่าแบบช่วง ผลการวิจัยสรุปได้ว่า (1) กรณีการประมาณค่าแบบจุด พบว่าโดยส่วนใหญ่ประมาณ 71.43% วิธีบูตสเตรป จะมีประสิทธิภาพดีกว่า วิธีแจ็กไนฟ์ (2) กรณีการประมาณค่าแบบช่วง พบว่า ทุกค่าพารามิเตอร์ ในทุกๆลักษณะการแจกแจง วิธีบูตสเตรป จะมีประสิทธิภาพดีกว่า วิธีแจ็กไนฟ์

ศรสวรรค์ บุญเพ็ง บุญอ้อม โหมที และอภิญญา หิรัญวงษ์ (2558) ศึกษาเปรียบเทียบวิธีการประมาณค่าแบบช่วงสำหรับพารามิเตอร์ขนาดของการแจกแจงไวบูลล์แบบสองพารามิเตอร์ โดยมีวัตถุประสงค์ของงานวิจัยคือศึกษาและเปรียบเทียบวิธีการประมาณค่าแบบช่วงสำหรับพารามิเตอร์ขนาด (η) ของการแจกแจงแบบไวบูลล์แบบสองพารามิเตอร์ เมื่อทราบค่าพารามิเตอร์ (β) ด้วยวิธีการประมาณค่า 2 วิธีคือวิธีบูตสเตรปที และวิธีแจ็กไนฟ์ โดยพิจารณาค่าสัมประสิทธิ์ความเชื่อมั่น และค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่น เป็นเกณฑ์ในการเปรียบเทียบ วิธีที่ดีกว่าเป็นวิธีที่มีค่าสัมประสิทธิ์ความเชื่อมั่นไม่ต่ำกว่า ค่าประสิทธิ์ความเชื่อมั่น ไม่ต่ำกว่า ค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด และมีค่าความกว้างเฉลี่ยของความเชื่อมั่นต่ำที่สุด เมื่อกำหนดให้ข้อมูลขนาดตัวอย่าง (n) เท่ากับ 10, 20, 30, 50 และ 100 ค่าพารามิเตอร์ (β) มีค่าเท่ากับ 1, 2, 3, 4, 5 และ 6

สำหรับค่าพารามิเตอร์ (η) มีค่า 1, 1.5 , 2, 2.5 และ 3 โดยจำนวนครั้งที่สุ่มซ้ำด้วยวิธีบูตสเตรปที่เท่ากับ 1,000 ครั้ง และกำหนดสัมประสิทธิ์ความเชื่อมั่นที่ 95% ทำการจำลองข้อมูลโดยใช้เทคนิคมอนติคาร์โล ด้วยโปรแกรม SAS ทำการทดลองซ้ำ 1,000 ครั้ง ในแต่ละสถานการณ์ที่ศึกษา ผลการศึกษาพบว่า วิธีบูตสเตรปที่ เป็นวิธีที่ดีที่สุดสำหรับทุกๆ สถานการณ์ที่กำหนด โดยเฉพาะในกรณีที่ขนาดตัวอย่างเท่ากับ 50 และ 100 และค่าพารามิเตอร์ η ที่มีขนาดเล็กที่สุด ($\eta=1$) กับค่าพารามิเตอร์ที่มีขนาดใหญ่

จิราภรณ์ สิงห์ผง (2560) ศึกษาเกี่ยวกับการประมาณค่าช่วงความเชื่อมั่นของพารามิเตอร์ของการแจกแจงค่าสุดขีด เมื่อข้อมูลถูกเซ็นเซอร์ประเภทที่ 1 โดยวัตถุประสงค์ของการวิจัยคือศึกษาข้อมูลระยะเวลาการใช้งานที่มีการแจกแจงค่าสุดขีด และเป็นข้อมูลที่ถูกระบุเซ็นเซอร์ประเภทที่ 1 รวมถึงการประมาณค่าพารามิเตอร์ด้วยวิธีภาวะน่าจะเป็นสูงสุดและนำตัวประมาณภาวะน่าจะเป็นสูงสุดมาสร้างช่วงความเชื่อมั่นของพารามิเตอร์โดยใช้ เมทริกซ์ข้อสนเทศค่าสังเกตของฟิชเชอร์ในการสร้างช่วงความเชื่อมั่นซึ่งอาศัยวิธีเชิงเส้นกำกับปกติ และสร้างช่วงความเชื่อมั่นด้วยวิธีบูตสเตรปแบบใช้พารามิเตอร์ 2 วิธีที่ต่างกัน ได้แก่ วิธีบูตสเตรปที่ และวิธีบูตสเตรปที่ นอกจากนี้ยังใช้วิธีอัตราส่วนภาวะน่าจะเป็น และเปรียบเทียบประสิทธิภาพ จากทั้ง 4 วิธี ด้วยค่าความน่าจะเป็นกลุ่มรวมและค่าเฉลี่ยความกว้างของช่วงความเชื่อมั่น จากการ จำลองข้อมูล รวมถึงนำวิธีที่นำเสนอมาประมาณค่าช่วงความเชื่อมั่นของพารามิเตอร์โดยใช้ข้อมูลจริง

บทที่ 3

วิธีดำเนินงานวิจัย

การวิจัยในครั้งนี้ศึกษาเกี่ยวกับช่วงความเชื่อมั่นสำหรับค่าเฉลี่ยประชากรภายใต้การสุ่มโดยใช้ความน่าจะเป็นในการเลือกไม่เท่ากัน กรณีที่เป็นการสุ่มแบบไม่ใส่คืนค่าเฉลี่ยประชากร สามารถประมาณค่าได้จากช่วงความเชื่อมั่นของ Horvitz และ Thompson (1952) และตัวประมาณค่าของ Hajek (1971) อย่างไรก็ตามช่วงความเชื่อมั่นดังกล่าวค่อนข้างยุ่งยากในทางปฏิบัติเนื่องจากความซับซ้อนในการคำนวณค่า π_{ij} ที่มีค่าแตกต่างกันในแต่ละหน่วย i และ j ดังนั้นการวิจัยในครั้งนี้ผู้วิจัยต้องการพัฒนาช่วงความเชื่อมั่นที่ได้กล่าวถึงข้างต้นโดยไม่จำเป็นต้องใช้ค่า π_{ij} ในการคำนวณค่า และเรียกช่วงความเชื่อมั่นที่ได้ว่าช่วงความเชื่อมั่นแบบเชิงเส้น และช่วงความเชื่อมั่นแบบไม่เป็นเชิงเส้น ผู้วิจัยกล่าวถึงสัญลักษณ์และข้อตกลงเบื้องต้นในหัวข้อ 3.1 หัวข้อ 3.2 กล่าวถึงวิธีการนำเสนอช่วงความเชื่อมั่นสำหรับการประมาณค่าเฉลี่ยประชากร สำหรับการเปรียบเทียบประสิทธิภาพช่วงความเชื่อมั่นที่นำเสนอ ถูกกล่าวถึงในหัวข้อ 3.3 และหัวข้อ 3.4 ได้อธิบายถึงเกณฑ์ที่ใช้ในการเปรียบเทียบประสิทธิภาพของช่วงความเชื่อมั่นที่นำเสนอ

3.1 สัญลักษณ์และข้อตกลงเบื้องต้น

สมมติว่าประชากรที่สนใจศึกษาประกอบด้วยหน่วยตัวอย่างทั้งหมด N หน่วย เขียนแทนด้วยสัญลักษณ์ $U = \{1, 2, \dots, N\}$ และตัวอย่าง s ขนาด n ถูกสุ่มจากประชากร U ด้วยแผนการสุ่มตัวอย่างแบบใช้ความน่าจะเป็นในการเลือกไม่เท่ากัน และไม่ใส่คืน กำหนดให้ y แทนตัวแปรที่สนใจศึกษา ส่วน k แทนตัวแปรช่วยที่มีความสัมพันธ์กับตัวแปร y กำหนดสัญลักษณ์ $\bar{Y} = N^{-1} \sum_{i \in U} y_i$ แทนค่าเฉลี่ยของประชากร (พารามิเตอร์ที่ต้องการศึกษา)

กำหนด $\pi_i = P(i \in s)$ แทนความน่าจะเป็นที่หน่วยตัวอย่างที่ i ถูกเลือกเป็นสมาชิกในตัวอย่าง s ขนาด n ส่วน $\pi_{ij} = P(i, j \in s)$ แทนความน่าจะเป็นที่หน่วยตัวอย่างที่ i และ j เป็นสมาชิกของตัวอย่าง s ขนาด n ภายใต้แผนการสุ่มตัวอย่างแบบง่ายและไม่ใส่คืนค่า π_i และ π_{ij} มีค่าคงที่กล่าวคือ $\pi_i = nN^{-1}$ และ $\pi_{ij} = n(n-1)(N(N-1))^{-1}$ แต่แผนการสุ่มตัวอย่างแบบใช้ความน่าจะเป็นในการเลือกไม่เท่ากันและไม่ใส่คืน ค่า π_i และ π_{ij} มีค่าแตกต่างกันตามค่าของ k_i หรือ k_j ส่งผลให้ $C.I._{(1)}(\bar{Y})$ และ $C.I._{(2)}(\bar{Y})$ ค่อนข้างยุ่งยากในการคำนวณค่า เมื่อ $C.I._{(1)}(\bar{Y})$ และ $C.I._{(2)}(\bar{Y})$ แทนช่วงความเชื่อมั่นที่สร้างจากตัวประมาณค่าของ Horvitz และ Thompson (1952) และ ตัวประมาณค่าของ Hajek (1971) ตามลำดับ ดังนั้นการวิจัยในครั้งนี้ผู้วิจัยนำเสนอช่วงความเชื่อมั่นสำหรับประมาณค่าเฉลี่ยประชากร ที่พัฒนาจาก $C.I._{(1)}(\bar{Y})$ และ $C.I._{(2)}(\bar{Y})$ โดยไม่จำเป็นต้องใช้ค่า π_{ij} ในการคำนวณค่า

3.2 การนำเสนอช่วงความเชื่อมั่นสำหรับการประมาณค่าเฉลี่ยประชากร

3.2.1 ช่วงความเชื่อมั่นแบบเป็นเชิงเส้น

การนำเสนอช่วงความเชื่อมั่นแบบเป็นเชิงเส้น เขียนแทนด้วยสัญลักษณ์ $Adjusted\ C.I._{(1)}(\bar{Y})$ ผู้วิจัยพัฒนาช่วงความเชื่อมั่นที่นำเสนอ จากการปรับช่วงความเชื่อมั่น $C.I._{(1)}(\bar{Y})$ ที่สร้างจากตัวประมาณของ Horvitz และ Thompson (1952) โดยการหาตัวประมาณค่าความแปรปรวนของ $\hat{Y}_{HT}$ ด้วยวิธีของ Hansen และ Horvitz (1946) เขียนแทนด้วย $\hat{V}_1$ จากนั้นสร้างตัวแปรสุ่ม $Z_1 = \frac{\hat{Y}_{HT} - \bar{Y}}{\sqrt{\hat{V}_1}} = \frac{\hat{Y}_1 - \bar{Y}}{\sqrt{\hat{V}_1}}$ เมื่อ $\hat{Y}_1 = \hat{Y}_{HT}$ และอาศัยทฤษฎีเข้าสู่ส่วนกลางเพื่อแสดงว่า $Z_1 \sim appN(0,1)$ สุดท้ายผู้วิจัย สร้างช่วงความเชื่อมั่น $Adjusted\ C.I._{(1)}(\bar{Y})$ ตามวิธีของ กัลยา วาณิชย์บัญชา (2547) และ Casella และ Berger (2002)

3.2.2 ช่วงความเชื่อมั่นแบบไม่เป็นเชิงเส้น

การนำเสนอช่วงความเชื่อมั่นแบบไม่เป็นเชิงเส้นเขียนแทนด้วยสัญลักษณ์ $Adjusted\ C.I._{(2)}(\bar{Y})$ ผู้วิจัยใช้แนวคิดเดียวกับการสร้าง $Adjusted\ C.I._{(1)}(\bar{Y})$ โดยการหาตัวประมาณค่าความแปรปรวนของตัวประมาณค่าแบบจุดของ Hajek (1971) หรือ $\hat{Y}_{Hajek}$ ด้วยวิธีของ Hansen และ Horvitz (1946) เขียนแทนด้วย $\hat{V}_2$ จากนั้นสร้างตัวแปรสุ่ม $Z_2 = \frac{\hat{Y}_{Hajek} - \bar{Y}}{\sqrt{\hat{V}_2}} = \frac{\hat{Y}_2 - \bar{Y}}{\sqrt{\hat{V}_2}}$ เมื่อ $\hat{Y}_2 = \hat{Y}_{Hajek}$ และอาศัยทฤษฎีเข้าสู่ส่วนกลางเพื่อแสดงว่า $Z_2 \sim appN(0,1)$ สุดท้ายผู้วิจัย สร้างช่วงความเชื่อมั่น $Adjusted\ C.I._{(2)}(\bar{Y})$ ตามวิธีของ กัลยา วาณิชย์บัญชา (2547) และ Casella และ Berger (2002)

3.3 การเปรียบเทียบประสิทธิภาพช่วงความเชื่อมั่นที่นำเสนอ

ผู้วิจัยเปรียบเทียบประสิทธิภาพช่วงความเชื่อมั่นที่นำเสนอ โดยใช้เทคนิคมอนติคาร์โลซิมูเลชันและกำหนดการทำซ้ำทั้งหมด 50,000 รอบตามขั้นตอนต่อไปนี้

3.3.1 ประชากรที่ใช้ในการศึกษา

การวิจัยในครั้งนี้ศึกษาเกี่ยวกับการประมาณค่าเฉลี่ยประชากรของตัวแปรที่สนใจศึกษา โดยใช้ตัวประมาณค่าแบบช่วง และสุ่มตัวอย่าง s ขนาด n จากประชากร U ขนาด N โดยใช้แผนการสุ่มตัวอย่างแบบใช้ความน่าจะเป็นในการเลือกไม่เท่ากันและไม่ใส่คืน ซึ่งประชากรที่ใช้ในการเปรียบเทียบประสิทธิภาพของช่วงความเชื่อมั่น ผู้วิจัยทำการจำลองข้อมูล $N=10,000$ และ

พิจารณากรณีที่ y_i มีความสัมพันธ์ในทิศทางตรงข้ามกับ π_i เมื่อ $i \in U$ และ $U = \{1, 2, 3, \dots, 10,000\}$ ตามการศึกษาของ Tillé และ Matei (2016) ดังนี้

- (1) กำหนดค่าของตัวแปรช่วย $k_i = i$ เมื่อ $i \in U$
- (2) ใช้ฟังก์ชัน inclusion probabilities () ใน package “Sampling” ของโปรแกรม R เพื่อกำหนดค่า π_i เมื่อ $i \in U$
- (3) กำหนดค่าตัวแปรที่สนใจศึกษา $y_i = \pi_i^{-1}$ เมื่อ $i \in U$

3.3.2 คำนวณความกว้างเฉลี่ยของช่วงความเชื่อมั่น

การคำนวณความกว้างเฉลี่ยของช่วงความเชื่อมั่นผู้วิจัยกำหนดการซ้ำทั้งหมด 50,000 รอบ และ $M = \{1, 2, \dots, 50,000\}$ และกำหนดให้

$L_1[m]$ และ $U_1[m]$ แทนขีดความเชื่อมั่นล่าง และขีดความเชื่อมั่นบน ของช่วงความเชื่อมั่นที่เป็นเชิงเส้น

$L_2[m]$ และ $U_2[m]$ แทนขีดความเชื่อมั่นล่าง และขีดความเชื่อมั่นบน ของช่วงความเชื่อมั่นที่ไม่เป็นเชิงเส้น

- (i) กำหนดระดับความเชื่อมั่น $(1-\alpha)100\% = 90\%$
- (ii) สำหรับทุก $n \in M$ สุ่มตัวอย่าง s ขนาด $n=30$ จากประชากร U และคำนวณค่า $L_k[m]$ และ $U_k[m]$ เมื่อ $k=1, 2$
- (iii) คำนวณค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่น เขียนแทนด้วยค่า AL และมีค่าตามที่กำหนดในสมการ (3-1)

$$AL_k = \frac{1}{M} \sum_{m=1}^M (U_k[m] - L_k[m]) \text{ เมื่อ } k=1, 2 \quad (3-1)$$

- (iv) ทำซ้ำขั้นตอนที่ (i) ถึง (iii) แต่เปลี่ยนขนาดตัวอย่างจาก $n=30$ เป็น $n=35, 40, 45$ และ 50 ตามลำดับ

- (v) ทำซ้ำขั้นตอนที่ (i) ถึง (iv) แต่เปลี่ยนระดับความเชื่อมั่นจาก 90% เป็น 95% และ 99% ตามลำดับ

3.4 เสนอการเปรียบเทียบประสิทธิภาพช่วงความเชื่อมั่น

สำหรับเสนอการเปรียบเทียบประสิทธิภาพช่วงความเชื่อมั่นที่นำเสนอผู้วิจัยใช้เกณฑ์ตามที่กล่าวถึงในบทนิยาม 3.1

บทนิยาม 3.1 กำหนดให้ $C.I._A(\bar{Y})$ และ $C.I._B(\bar{Y})$ แทนช่วงความเชื่อมั่นสำหรับประมาณค่าเฉลี่ยประชากร จะกล่าวว่า $C.I._A(\bar{Y})$ มีประสิทธิภาพดีกว่า $C.I._B(\bar{Y})$ ก็ต่อเมื่อ $AL_A < AL_B$ เมื่อ AL_A และ AL_B แทนความกว้างเฉลี่ยของ $C.I._A(\bar{Y})$ และ $C.I._B(\bar{Y})$ ตามลำดับ

บทที่ 4

ผลการวิจัย

บทที่ 3 ผู้วิจัยได้กล่าวถึงวิธีการดำเนินการวิจัย สำหรับการสร้างช่วงความเชื่อมั่นเพื่อใช้ประมาณค่าเฉลี่ยประชากรภายใต้แผนการสุ่มตัวอย่างแบบใช้ความน่าจะเป็นในการเลือกไม่เท่ากัน และไม่ใส่คืน โดยผู้วิจัยพิจารณาช่วงความเชื่อมั่นที่น่าเสนอ เฉพาะกรณีที่สัดส่วนตัวอย่างมีขนาดเล็ก สำหรับบทนี้เป็นการนำเสนอผลการสร้างช่วงความเชื่อมั่น และผลการเปรียบเทียบประสิทธิภาพของช่วงความเชื่อมั่นที่น่าเสนอดังนี้

4.1 ช่วงความเชื่อมั่นสำหรับประมาณค่าเฉลี่ยประชากรที่น่าเสนอ

4.1.1 ช่วงความเชื่อมั่นแบบเป็นเชิงเส้น

หัวข้อนี้ผู้วิจัยจะได้กล่าวถึงช่วงความเชื่อมั่นแบบเป็นเชิงเส้น เพื่อใช้ประมาณค่าเฉลี่ยประชากร โดยช่วงความเชื่อมั่นที่ผู้วิจัยนำเสนอได้พัฒนามาจากช่วงความเชื่อมั่นที่สร้างจากตัวประมาณค่าของ Horvitz และ Thompson (1952) ดังนี้

จากสมการ (1-8) สามารถเขียนตัวประมาณค่าของ Horvitz และ Thompson (1952) ให้อยู่ในรูป

$$\hat{Y}_{HT} = \hat{Y}_1 = \frac{1}{N} \sum_{i \in S} \frac{y_i}{\pi_i} = \frac{1}{N} \sum_{i \in S} b_i \quad (4-1)$$

เมื่อ $b_i = f(y_i) = y_i \pi_i^{-1}$ จากนั้นหาตัวประมาณค่าความแปรปรวนของ $\hat{Y}_1$ ด้วยวิธีของ Hansen และ Horvitz (1946) เขียนแทนด้วยสัญลักษณ์ $\hat{V}_1$ เนื่องจากค่า $b_i = y_i \pi_i^{-1}$ ไม่ปรากฏตัวพารามิเตอร์ที่ไม่ทราบค่า ดังนั้น $\hat{b}_i = b_i = y_i \pi_i^{-1}$ และแทนค่าในสมการที่ (1-7) ใน บทที่ 1 ได้ว่า

$$\begin{aligned} \hat{V}_1 &= \frac{1}{N^2} \left(\frac{n}{n-1} \right) \left[\sum_{i \in S} \left(\frac{y_i}{\pi_i} \right)^2 - \frac{1}{n} \left(\sum_{i \in S} \frac{y_i}{\pi_i} \right)^2 \right] \\ &= \frac{1}{N^2} \left(\frac{n}{n-1} \right) \left[\sum_{i \in S} \bar{y}_{li}^2 - \frac{1}{n} \left(\sum_{i \in S} \bar{y}_{li} \right)^2 \right] \end{aligned} \quad (4-2)$$

เมื่อ $\bar{y}_{li} = y_i \pi_i^{-1}$ จากสมการที่ (4-1) และ (4-2) กำหนดตัวแปรสุ่ม $Z_1 = (\hat{Y}_1 - \bar{Y}) / \sqrt{\hat{V}_1}$ กรณีที่ตัวอย่างมีขนาดใหญ่ และใช้ทฤษฎีเข้าสู่ส่วนกลาง สามารถแสดงได้ว่า

$$Z_1 = \frac{(\hat{Y}_1 - \bar{Y})}{\sqrt{\hat{V}_1}} \sim \text{app}N(0,1) \quad (4-3)$$

จากนั้นใช้วิธีของ กัลยา วาณิชย์บัญชา (2547) และ Casella และ Berger (2002) สร้างช่วงความเชื่อมั่นแบบเชิงเส้น และมีค่าตามที่แสดงในทฤษฎี 4.1

ทฤษฎี 4.1 กำหนดให้ y แทนตัวแปรที่สนใจศึกษา และ $\bar{Y}$ แทนค่าเฉลี่ยประชากร สมมติว่าตัวอย่าง s ขนาด n ถูกสุ่มจากประชากร U ขนาด N โดยแผนการสุ่มตัวอย่างแบบใช้ความน่าจะเป็นในการเลือกไม่เท่ากันและไม่ใส่คืน และ n มีขนาดใหญ่รวมถึงไม่มีการสูญหายของข้อมูล กำหนดค่า $\hat{Y}_1$ ตามที่แสดงในสมการที่ (4-1) และ $\hat{V}_1$ ตามที่แสดงในสมการที่ (4-2) ช่วงความเชื่อมั่นแบบเชิงเส้นเพื่อใช้ประมาณค่า $\bar{Y}$ ที่ระดับ $(1-\alpha)100\%$ เขียนแทนด้วย $Adjusted\ C.I._{(1)}(\bar{Y})$ มีค่าตามที่แสดงในสมการที่ (4-4)

$$Adjusted\ C.I._{(1)}(\bar{Y}) = \left[\hat{Y}_1 - z_{1-\frac{\alpha}{2}} \sqrt{\hat{V}_1}, \hat{Y}_1 + z_{1-\frac{\alpha}{2}} \sqrt{\hat{V}_1} \right] = [L_1, U_1] \quad (4-4)$$

$$\text{เมื่อ } L_1 = \hat{Y}_1 - z_{1-\frac{\alpha}{2}} \sqrt{\hat{V}_1} \text{ และ } U_1 = \hat{Y}_1 + z_{1-\frac{\alpha}{2}} \sqrt{\hat{V}_1}$$

พิสูจน์ กำหนดค่า $\hat{Y}_1$ และ $\hat{V}_1$ ตามที่แสดงในสมการที่ (4-1) และ (4-2) ตามลำดับ จากสมการที่

$$(4-3) \text{ กรณีที่ } n \text{ มีขนาดใหญ่ } Z_1 = \frac{(\hat{Y}_1 - \bar{Y})}{\sqrt{\hat{V}_1}} \sim appN(0,1) \text{ เมื่อใช้วิธีของ กัลยา วาณิชย์บัญชา (2547)}$$

และ Casella และ Berger (2002) ได้ว่า

$$\begin{aligned} P\left(z_{\frac{\alpha}{2}} < Z_1 < z_{1-\frac{\alpha}{2}}\right) &= 1-\alpha \\ P\left(z_{\frac{\alpha}{2}} < \frac{\hat{Y}_1 - \bar{Y}}{\sqrt{\hat{V}_1}} < z_{1-\frac{\alpha}{2}}\right) &= 1-\alpha \\ P\left(-z_{1-\frac{\alpha}{2}} < \frac{\hat{Y}_1 - \bar{Y}}{\sqrt{\hat{V}_1}} < z_{1-\frac{\alpha}{2}}\right) &= 1-\alpha \end{aligned}$$

ดังนั้นช่วงความเชื่อมั่นแบบของ $\bar{Y}$ ที่ระดับ $(1-\alpha)100\%$ มีค่าเป็น

$$Adjusted\ C.I._{(1)}(\bar{Y}) = \left[\hat{Y}_1 - z_{1-\frac{\alpha}{2}} \sqrt{\hat{V}_1}, \hat{Y}_1 + z_{1-\frac{\alpha}{2}} \sqrt{\hat{V}_1} \right] = [L_1, U_1]$$

$$\text{เมื่อ } L_1 = \hat{Y}_1 - z_{1-\frac{\alpha}{2}} \sqrt{\hat{V}_1} \text{ และ } U_1 = \hat{Y}_1 + z_{1-\frac{\alpha}{2}} \sqrt{\hat{V}_1}$$

4.1.2 ช่วงความเชื่อมั่นแบบไม่เป็นเชิงเส้น

หัวข้อ 4.1.1 ผู้วิจัยได้นำเสนอช่วงความเชื่อมั่นแบบเชิงเส้น เพื่อใช้ประมาณค่า $\bar{Y}$ โดยช่วงความเชื่อมั่นที่ผู้วิจัยนำเสนอพัฒนามาจากช่วงความเชื่อมั่นที่สร้างจากตัวประมาณค่าของ Horvitz และ Thompson (1952) สำหรับหัวข้อนี้ผู้วิจัยจะได้นำเสนอช่วงความเชื่อมั่นที่พัฒนามาจากตัวประมาณค่าของ Hájek (1971) และเรียกช่วงความเชื่อมั่นดังกล่าวว่าช่วงความเชื่อมั่นแบบไม่เป็นเชิงเส้นดังนี้ จากสมการที่ (1-11) Hájek (1971) ได้นำเสนอตัวประมาณค่าแบบไม่เป็นเชิงเส้น เพื่อใช้ประมาณค่า $\bar{Y}$ ตามที่แสดงในสมการที่ (4-5)

$$\hat{\bar{Y}}_{Ha'jek} = \sum_{i \in S} \frac{y_i}{\pi_i} / \sum_{i \in S} \frac{1}{\pi_i} \quad (4-5)$$

อย่างไรก็ตาม $\hat{\bar{Y}}_{Ha'jek}$ ตามที่แสดงในสมการ (4-5) เป็นตัวประมาณค่าแบบไม่เป็นเชิงเส้น ดังนั้นการหาตัวประมาณค่าความแปรปรวน จำเป็นต้องใช้เทเลอร์แบบเชิงเส้นในการปรับตัวประมาณค่าดังกล่าวให้เป็นเชิงเส้นก่อนหาตัวประมาณค่าความแปรปรวน และเมื่อใช้วิธีเทเลอร์แบบเชิงเส้น สามารถเขียน $\hat{\bar{Y}}_{Ha'jek}$ ให้อยู่ในรูป

$$\begin{aligned} \hat{\bar{Y}}_{Ha'jek} &= \hat{\bar{Y}}_2 \\ &\approx \frac{1}{N} \sum_{i \in S} \frac{1}{\pi_i} (y_i - \bar{Y}) = \frac{1}{N} \sum_{i \in S} b_i \end{aligned} \quad (4-6)$$

เมื่อ $b_i = \frac{1}{\pi_i} (y_i - \bar{Y})$ จากนั้นหาตัวประมาณค่าความแปรปรวนของ $\hat{\bar{Y}}_{Ha'jek}$ ด้วยวิธีของ Hansen และ Horvitz (1946) เขียนแทนด้วยสัญลักษณ์ $\hat{V}_2$ อย่างไรก็ตาม ค่า $b_i = \frac{1}{\pi_i} (y_i - \bar{Y})$ มีพารามิเตอร์ที่ไม่ทราบค่าคือ $\bar{Y}$ และสามารถประมาณค่าด้วย $\hat{\bar{Y}}_{Ha'jek}$ ดังนั้น $\hat{b}_i = \frac{1}{\pi_i} (y_i - \hat{\bar{Y}}_{Ha'jek})$ และแทนค่า $\hat{b}_i$ ในสมการที่ (1-7) ตามที่ได้กล่าวถึงในบทที่ 1 ได้ว่า

$$\begin{aligned} \hat{V}_2 &= \frac{1}{N^2} \left(\frac{n}{n-1} \right) \left[\sum_{i \in S} \left(\frac{1}{\pi_i} (y_i - \hat{\bar{Y}}_{Ha'jek}) \right)^2 - \frac{1}{n} \left(\sum_{i \in S} \frac{1}{\pi_i} (y_i - \hat{\bar{Y}}_{Ha'jek}) \right)^2 \right] \\ &= \frac{1}{N^2} \left(\frac{n}{n-1} \right) \left[\sum_{i \in S} \bar{y}_{2i}^2 - \frac{1}{n} \left(\sum_{i \in S} \bar{y}_{2i} \right)^2 \right] \end{aligned} \quad (4-7)$$

เมื่อ $\bar{y}_{2i} = \pi_i^{-1} (y_i - \hat{\bar{Y}}_{Ha'jek})$ จากนั้นสร้างตัวแปรสุ่ม $Z_2 = (\hat{\bar{Y}}_2 - \bar{Y}) / \sqrt{\hat{V}_2}$ ซึ่งกรณีนี้ตัวอย่างมีขนาดใหญ่ และใช้ทฤษฎีเข้าสู่ส่วนกลาง สามารถแสดงได้ว่า

$$Z_2 = \frac{(\hat{\bar{Y}}_2 - \bar{Y})}{\sqrt{\hat{V}_2}} \sim appN(0,1) \quad (4-8)$$

ดังนั้นช่วงความเชื่อมั่นแบบไม่เป็นเชิงเส้น เพื่อใช้ประมาณค่าเฉลี่ยประชากร โดยใช้วิธีของ กัลยา วาณิชย์บัญชา (2547) และ Casella และ Berger (2002) มีค่าตามที่แสดงในทฤษฎี 4.2

ทฤษฎี 4.2 กำหนดให้ y แทนตัวแปรที่สนใจศึกษา และ $\bar{Y}$ แทนค่าเฉลี่ยประชากร สมมติว่าตัวอย่าง s ขนาด n ถูกสุ่มจากประชากร U ขนาด N โดยแผนการสุ่มตัวอย่างแบบใช้ความน่าจะเป็นในการเลือกไม่เท่ากันและไม่ใช่คืน และ n มีขนาดใหญ่รวมถึงไม่มีการสูญหายของข้อมูล กำหนดค่า $\hat{Y}_2$ ตามที่แสดงในสมการที่ (4-6) และ $\hat{V}_2$ ตามที่แสดงในสมการที่ (4-7) ช่วงความเชื่อมั่นแบบไม่เป็นเชิงเส้นของ $\bar{Y}$ ที่ระดับ $(1-\alpha)100\%$ เขียนแทนด้วย $Adjusted\ C.I._{(2)}(\bar{Y})$ และกำหนดค่าตามที่แสดงในสมการที่ (4-8)

$$Adjusted\ C.I._{(2)}(\bar{Y}) = \left[\hat{Y}_2 - z_{1-\frac{\alpha}{2}} \sqrt{\hat{V}_2}, \hat{Y}_2 + z_{1-\frac{\alpha}{2}} \sqrt{\hat{V}_2} \right] = [L_2, U_2] \quad (4-8)$$

$$\text{เมื่อ } L_2 = \hat{Y}_2 - z_{1-\frac{\alpha}{2}} \sqrt{\hat{V}_2} \text{ และ } U_2 = \hat{Y}_2 + z_{1-\frac{\alpha}{2}} \sqrt{\hat{V}_2}$$

การพิสูจน์ เช่นเดียวกับทฤษฎี 4.1

4.1.3 สูตรทั่วไปสำหรับช่วงความเชื่อมั่นที่น่าเสนอ

หัวข้อที่ 4.1.1 และ 4.1.2 ผู้วิจัยได้นำเสนอช่วงความเชื่อมั่นแบบเป็นเชิงเส้นและช่วงความเชื่อมั่นแบบไม่เป็นเชิงเส้นเพื่อใช้ประมาณค่าเฉลี่ยประชากร ในกรณีที่สัดส่วนตัวอย่างมีขนาดเล็กและขนาดตัวอย่างมีขนาดใหญ่ สำหรับข้อนี้ผู้วิจัยจะได้นำเสนอสูตรทั่วไปของช่วงความเชื่อมั่นที่น่าเสนอ ตามที่แสดงในบทตั้ง 4.3

บทตั้ง 4.3 กำหนดให้ y แทนตัวแปรที่สนใจศึกษา และ $\bar{Y}$ แทนค่าเฉลี่ยประชากร สมมติว่าตัวอย่าง s ขนาด n ถูกสุ่มจากประชากร U ขนาด N โดยแผนการสุ่มตัวอย่างแบบใช้ความน่าจะเป็นในการเลือกไม่เท่ากันและไม่ใช่คืน และ n มีขนาดใหญ่รวมถึงไม่มีการสูญหายของข้อมูลของตัวแปรที่สนใจศึกษาและตัวแปรช่วย ช่วงความเชื่อมั่นของ $\bar{Y}$ ที่ระดับ $(1-\alpha)100\%$ เขียนแทนด้วย $Adjusted\ C.I._{(k)}(\bar{Y})$ เมื่อ $k=1,2$ กำหนดค่าตามที่แสดงในสมการที่ (4-9)

$$Adjusted\ C.I._{(k)}(\bar{Y}) = \left[\hat{Y}_k - z_{1-\frac{\alpha}{2}} \sqrt{\hat{V}_k}, \hat{Y}_k + z_{1-\frac{\alpha}{2}} \sqrt{\hat{V}_k} \right] = [L_k, U_k] \text{ เมื่อ } k=1,2 \quad (4-9)$$

$$\text{เมื่อ } \hat{Y}_1 = \frac{1}{N} \sum_{i \in s} \frac{y_i}{\pi_i}$$

$$\hat{V}_1 = \frac{1}{N^2} \left(\frac{n}{n-1} \right) \left[\sum_{i \in s} \bar{y}_{1i}^2 - \frac{1}{n} \left(\sum_{i \in s} \bar{y}_{1i} \right)^2 \right]$$

$$\hat{Y}_2 = \hat{Y}_{Ha'jek} = \sum_{i \in s} \frac{y_i}{\pi_i} / \sum_{i \in s} \frac{1}{\pi_i}$$

$$\hat{V}_2 = \frac{1}{N^2} \left(\frac{n}{n-1} \right) \left[\sum_{i \in S} \tilde{y}_{2i}^2 - \frac{1}{n} \left(\sum_{i \in S} \tilde{y}_{2i} \right)^2 \right]$$

$$\tilde{y}_{li} = y_i \pi_i^{-1}$$

$$\tilde{y}_{2i} = \pi_i^{-1} \left(y_i - \hat{Y}_{Ha'jek} \right)$$

$$L_1 = \hat{Y}_1 - z_{1-\frac{\alpha}{2}} \sqrt{\hat{V}_1} \text{ และ } U_1 = \hat{Y}_1 + z_{1-\frac{\alpha}{2}} \sqrt{\hat{V}_1}$$

$$L_2 = \hat{Y}_2 - z_{1-\frac{\alpha}{2}} \sqrt{\hat{V}_2} \text{ และ } U_2 = \hat{Y}_2 + z_{1-\frac{\alpha}{2}} \sqrt{\hat{V}_2}$$

4.2 ผลการเปรียบเทียบช่วงความเชื่อมั่นสำหรับประมาณค่าเฉลี่ยประชากรที่นำเสนอ

หัวข้อ 4.1 ผู้วิจัยได้นำเสนอช่วงความเชื่อมั่นแบบเชิงเส้น และช่วงความเชื่อมั่นแบบไม่เป็นเชิงเส้นเพื่อใช้ประมาณค่าเฉลี่ยประชากร รวมถึงนำเสนอสูตรทั่วไปของช่วงความเชื่อมั่นดังกล่าว สำหรับหัวข้อนี้เป็นการนำเสนอค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่น ที่ระดับความเชื่อมั่น 90% 95% และ 99% เพื่อใช้เปรียบเทียบประสิทธิภาพของช่วงความเชื่อมั่นที่นำเสนอ ตามที่แสดงในตารางที่ 4.1 ตารางที่ 4.2 และตารางที่ 4.3 ตามลำดับ

ผลจากตารางที่ 4.1 - ตารางที่ 4.3 พบว่าในแต่ละระดับของความเชื่อมั่น ค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่นที่ไม่เป็นเชิงเส้นมีค่าน้อยกว่าช่วงความเชื่อมั่นที่เป็นเชิงเส้น ทุกระดับของขนาดตัวอย่าง ดังนั้นโดยอาศัยบทนิยาม 3.1 ตามที่ได้กล่าวในบทที่ 3 จะได้ว่าช่วงความเชื่อมั่นแบบไม่เป็นเชิงเส้นมีประสิทธิภาพดีกว่าช่วงความเชื่อมั่นแบบเชิงเส้นเชิงเส้น

ตารางที่ 4.1 ค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่นแบบเชิงเส้น และช่วงความเชื่อมั่นแบบไม่เป็นเชิงเส้น ที่ระดับความเชื่อมั่น 90%

n	$Adjusted\ C.I._{(1)}(\bar{Y})$	$Adjusted\ C.I._{(2)}(\bar{Y})$
30	645.83	137.40
35	219.52	97.02
40	159.73	81.93
45	301.09	89.64
50	116.69	60.40

ตารางที่ 4.2 ค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่นแบบเชิงเส้น และช่วงความเชื่อมั่นแบบไม่เป็นเชิงเส้น ที่ระดับความเชื่อมั่น 95%

n	$Adjusted\ C.I._{(1)}(\bar{Y})$	$Adjusted\ C.I._{(2)}(\bar{Y})$
30	442.67	146.82
35	233.38	116.39
40	173.62	94.61
45	169.424	91.76
50	164.08	78.85

ตารางที่ 4.3 ค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่นแบบเชิงเส้น และช่วงความเชื่อมั่นแบบไม่เป็นเชิงเส้น ที่ระดับความเชื่อมั่น 99%

n	$Adjusted\ C.I._{(1)}(\bar{Y})$	$Adjusted\ C.I._{(2)}(\bar{Y})$
30	174.13	107.63
35	179.07	107.29
40	138.54	106.13
45	136.99	101.97
50	127.49	81.39

บทที่ 5

สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ

การวิจัยในครั้งนี้มีวัตถุประสงค์เพื่อนำเสนอ และเปรียบเทียบประสิทธิภาพของช่วงความเชื่อมั่นที่ใช้ประมาณค่าเฉลี่ยประชากร ภายใต้การสุ่มตัวอย่างแบบใช้ความน่าจะเป็นในการเลือกไม่เท่ากัน โดยศึกษาในกรณีที่เป็นการสุ่มแบบไม่ใส่คืน และสัดส่วนตัวอย่างมีขนาดเล็ก (ขนาดตัวอย่างไม่เกิน 5% ของขนาดประชากร) ช่วงความเชื่อมั่นที่ผู้วิจัยนำเสนอพัฒนามาจากช่วงความเชื่อมั่นที่สร้างจากตัวประมาณค่าของ Horvitz และ Thompson (1952) และ Hájek (1971) เรียกช่วงความเชื่อมั่นที่ได้ว่า ช่วงความเชื่อมั่นแบบเชิงเส้น และช่วงความเชื่อมั่นแบบไม่เป็นเชิงเส้นตามลำดับ ผู้วิจัยเปรียบเทียบประสิทธิภาพของช่วงความเชื่อมั่นที่นำเสนอด้วยค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่น และใช้โปรแกรม R ในการประมวลผล ซึ่งในบทนี้จะกล่าวถึงผลสรุปของการวิจัย อภิปรายผล และมีข้อเสนอแนะดังนี้

5.1 สรุปผลการวิจัย

5.1.1 ช่วงความเชื่อมั่นสำหรับประมาณค่าเฉลี่ยประชากร

สูตรทั่วไปของช่วงความเชื่อมั่นสำหรับประมาณค่าเฉลี่ยประชากร ที่ระดับความเชื่อมั่น $(1-\alpha)100\%$ ที่ผู้วิจัยนำเสนอแสดงดังสมการที่ (5-1)

$$\text{Adjusted } C.I._{(k)}(\bar{Y}) = \left[\hat{\bar{Y}}_k - z_{1-\frac{\alpha}{2}} \sqrt{\hat{V}_k}, \hat{\bar{Y}}_k + z_{1-\frac{\alpha}{2}} \sqrt{\hat{V}_k} \right] = [L_k, U_k] \text{ เมื่อ } k=1,2 \quad (5-1)$$

เมื่อ $\text{Adjusted } C.I._{(1)}(\bar{Y})$ แทนช่วงความเชื่อมั่นแบบเชิงเส้น

$\text{Adjusted } C.I._{(2)}(\bar{Y})$ แทนช่วงความเชื่อมั่นแบบไม่เป็นเชิงเส้น

$$\hat{\bar{Y}}_1 = \frac{1}{N} \sum_{i \in S} \frac{y_i}{\pi_i}$$

$$\hat{V}_1 = \frac{1}{N^2} \left(\frac{n}{n-1} \right) \left[\sum_{i \in S} \bar{y}_{1i}^2 - \frac{1}{n} \left(\sum_{i \in S} \bar{y}_{1i} \right)^2 \right]$$

$$\hat{\bar{Y}}_2 = \hat{\bar{Y}}_{Ha'jek} = \sum_{i \in S} \frac{y_i}{\pi_i} / \sum_{i \in S} \frac{1}{\pi_i}$$

$$\hat{V}_2 = \frac{1}{N^2} \left(\frac{n}{n-1} \right) \left[\sum_{i \in S} \bar{y}_{2i}^2 - \frac{1}{n} \left(\sum_{i \in S} \bar{y}_{2i} \right)^2 \right]$$

$$\bar{y}_{li} = y_i \pi_i^{-1}$$

$$\begin{aligned}\hat{y}_{2i} &= \pi_i^{-1} \left(y_i - \hat{Y}_{Ha'jek} \right) \\ L_1 &= \hat{Y}_1 - z_{1-\frac{\alpha}{2}} \sqrt{\hat{V}_1} \text{ และ } U_1 = \hat{Y}_1 + z_{1-\frac{\alpha}{2}} \sqrt{\hat{V}_1} \\ L_2 &= \hat{Y}_2 - z_{1-\frac{\alpha}{2}} \sqrt{\hat{V}_2} \text{ และ } U_2 = \hat{Y}_2 + z_{1-\frac{\alpha}{2}} \sqrt{\hat{V}_2}\end{aligned}$$

5.1.2 ผลการเปรียบเทียบประสิทธิภาพของช่วงความเชื่อมั่นที่นำเสนอ

ผลการเปรียบเทียบประสิทธิภาพช่วงความเชื่อมั่นที่นำเสนอ พบว่าช่วงความเชื่อมั่นแบบไม่เป็นเชิงเส้น มีประสิทธิภาพดีกว่าช่วงความเชื่อมั่นแบบเชิงเส้น เนื่องจากค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่นแบบไม่เป็นเชิงเส้นมีค่าน้อยกว่าความกว้างเฉลี่ยของช่วงความเชื่อมั่นแบบเชิงเส้นทุกระดับความเชื่อมั่นและขนาดตัวอย่าง

5.2 การอภิปรายผล

เกณฑ์ที่ใช้ในการเปรียบเทียบประสิทธิภาพของช่วงความเชื่อมั่น ผู้วิจัยพิจารณาจากค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่น ซึ่ง [จิราภรณ์ สิงสง \(2560\)](#) กล่าวว่าช่วงความเชื่อมั่นที่ดีต้องมีค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่นแคบที่สุด ผลการเปรียบเทียบช่วงความเชื่อมั่นที่นำเสนอ พบว่าช่วงความเชื่อมั่นแบบไม่เป็นเชิงเส้นมีประสิทธิภาพดีกว่าช่วงความเชื่อมั่นแบบเชิงเส้น เนื่องจากค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่นแบบไม่เป็นเชิงเส้นแคบกว่า (น้อยกว่า) ช่วงความเชื่อมั่นที่แบบเชิงเส้น นอกจากนี้จากสมการที่ (5-1) จะเห็นว่าค่าเฉลี่ยของความกว้างของช่วงความเชื่อมั่น จะมีค่าเท่ากับค่าเฉลี่ยของผลต่างระหว่างจิตความเชื่อมั่นบน และจิตความเชื่อมั่นล่าง ซึ่งค่าดังกล่าวมีค่าขึ้นอยู่กับค่าของตัวอย่างที่ถูกสุ่มจากประชากร สำหรับขั้นตอนการแสดงตัวอย่างเชิงตัวเลข ผู้วิจัยกำหนดข้อมูลประชากร กรณีที่ y_i มีความสัมพันธ์ในทิศทางตรงข้ามกับ π_i ซึ่ง [Tillé และ Matei \(2016\)](#) ได้แสดงให้เห็นว่าภายใต้เงื่อนไขดังกล่าว ตัวประมาณแบบจุดที่นำเสนอโดย [Hájek \(1971\)](#) จะมีประสิทธิภาพดีกว่าตัวประมาณค่าแบบจุดของ [Horvitz และ Thompson \(1952\)](#) ซึ่งสอดคล้องกับผลการศึกษาของผู้วิจัย ที่พบว่าช่วงความเชื่อมั่นที่พัฒนาจากตัวประมาณค่าของ [Hájek \(1971\)](#) (ช่วงความเชื่อมั่นแบบไม่เป็นเชิงเส้น) มีประสิทธิภาพดีกว่าช่วงความเชื่อมั่นที่พัฒนาจากตัวประมาณค่าของ [Horvitz และ Thompson \(1952\)](#) (ช่วงความเชื่อมั่นแบบเชิงเส้น)

5.3 ข้อเสนอแนะ

5.3.1 ข้อเสนอแนะในการนำไปใช้

ช่วงความเชื่อมั่นที่ผู้วิจัยนำเสนอเหมาะสำหรับข้อมูลที่ทราบข้อมูลของตัวแปรช่วยก่อนเริ่มสุ่มตัวอย่าง เช่นการประมาณค่าผลผลิตเฉลี่ยของข้าวโพดเลี้ยงสัตว์ของเกษตรกรในเขตอำเภอหล่มสัก

จังหวัดเพชรบูรณ์ ซึ่งผู้วิจัยอาจใช้ข้อมูลพื้นที่ที่ใช้ในการเพาะปลูกของเกษตรกรในเขตอำเภอหล่มสัก เป็นข้อมูลตัวแปรช่วยซึ่งข้อมูลดังกล่าว สามารถขอได้จากสำนักงานการเกษตรอำเภอหล่มสัก หรือสำนักงานสถิติจังหวัดเพชรบูรณ์ เป็นต้น

5.32. ข้อเสนอแนะในการทำวิจัยครั้งต่อไป

ผู้สนใจศึกษาอาจขยายวิธีการของผู้วิจัยไปยังกรณีทั่วไป และใช้ วิธีแจ๊คไนฟ์ (Jackknife method) หรือวิธีบูตสเตรป (วิธีบูตสเตรป) ในการประมาณค่าความแปรปรวน เป็นต้น

บรรณานุกรม

- กัญญ์พิษฐา พุทธะไชยทัศน์ (2552). การเปรียบเทียบการประมาณการแจกแจงข้อมูลสมมาตรและ
ไม่สมมาตรด้วยวิธีแฉกในพีกับวิธีบูสเตรป. วิทยานิพนธ์ปริญญาโท จุฬาลงกรณ์มหา
วิทยาลัย กรุงเทพฯ.
- กัลยา วานิชย์บัญชา (2547). **หลักสถิติ (พิมพ์ครั้งที่ 4)**. กรุงเทพฯ:จุฬาลงกรณ์มหาวิทยาลัย.
- จิราภรณ์ สิงห์. (2560). การประมาณค่าช่วงความเชื่อมั่นของพารามิเตอร์ของการแจกแจงค่าสุดขีด
เมื่อข้อมูลถูกเซ็นเซอร์ประเภทที่ 1. วิทยานิพนธ์ วิทยาศาสตร์มหาบัณฑิต (สถิติประยุกต์)
สาขาวิชาคณิตศาสตร์และสถิติ. คณะวิทยาศาสตร์และเทคโนโลยี
มหาวิทยาลัยธรรมศาสตร์.
- ศรสวรรค์ บุญเพ็ง บุญอ้อม โฉมทิ และอภิญา หิรัญวงษ์ (2558). การเปรียบเทียบวิธีการประมาณ
ค่าแบบช่วงสำหรับพารามิเตอร์ขนาดของการแจกแจงไวบูลล์แบบสองพารามิเตอร์.
วารสารวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์. 23(4), หน้า 569-578.
- สุชาดา กิระนันท์. (2542). **ทฤษฎีและวิธีการสำรวจตัวอย่าง**. กรุงเทพฯ: โรงพิมพ์จุฬาลงกรณ์
มหาวิทยาลัย.
- Casella, G., and Berger., R.L. (2002). *Statistical inference*. Duxbury Press, second edition.
- Berger, Y. G. (2004). *A simple variance estimator for unequal probability sampling without
replacement. Journal of Applied Statistics.* Vol.31 : 305-315.
- Hájek, J. (1971). **Foundations of Statistical Inference**. Toronto, Canada: Holt, Rinehart,
Winston. Chap. Discussion of an essay on the logical foundations of survey sampling,
part on by D. Basu.
- Horvitz, D.G., and D.J. Thompson (1952). **A generalization of sampling without replacement
from a finite universe**. Journal of the American Statistical Association, 47, pp. 663-
685.
- Midzuno, H. (1952). **On sampling system with probability proportional to sum of sizes**, Ann.
Inst.Statist. Math. 3, pp 99-107.
- Särndal,C.E., Swenson, B. and Wretman, J.H. (1992). **Model assisted survey sampling**.
NewYork, Springer-Verlag.

Tillé, Y., Matei, M. (2016). *Sampling: Survey Sampling*. R package version 2.8. Accessed August 28, 2018. (2016), <https://CRAN.R-project.org/package=sampling>.

ภาคผนวก

ภาคผนวก ก

รายชื่อผู้ทรงคุณวุฒิตรวจสอบเครื่องมือ และเนื้อหาในโครงการวิจัย

รายชื่อผู้ทรงคุณวุฒิตรวจสอบเครื่องมือ และเนื้อหาในโครงการวิจัย

1. ผู้ช่วยศาสตราจารย์ อาดุลย์ จงรักษ์

ตำแหน่ง อาจารย์ประจำหลักสูตรวิทยาศาสตรบัณฑิต สาขาวิชาคณิตศาสตร์
คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏเพชรบูรณ์
จังหวัดเพชรบูรณ์

2. อาจารย์ หยาตพิรุณ สุภรากรสกุล

ตำแหน่ง อาจารย์ประจำหลักสูตรวิทยาศาสตรบัณฑิต สาขาวิชาคณิตศาสตร์
คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏเพชรบูรณ์
จังหวัดเพชรบูรณ์

ภาคผนวก ข
การเผยแพร่ผลงานวิจัย



การประชุมวิชาการระดับชาติ ครั้งที่ 3

รายงานสืบเนื่องการประชุมวิชาการ
และนำเสนอผลงานวิจัยระดับชาติ ครั้งที่ 3 ประจำปี 2562

The 3rd Roi Et Rajabhat University National and International
Conference on Education and Technology Research 2019
(RERU ICET III): Innovative Local Development

วันศุกร์ที่ 12 กรกฎาคม 2562
ณ หอประชุม 60 พรรษา มหาวิทยาลัยราชภัฏร้อยเอ็ด

15 กรกฎาคม 2562

บทความ	หน้า
การปรับปรุงกระบวนการปฏิบัติงานด้วยเทคนิคการลดความสูญเสียเปล่าแบบ ECRS กรณีศึกษา: แผนกจัดส่งสินค้า บริษัท เอปซี จำกัด วันพิชิต เบ้งจัน	1
การประมาณค่าแบบช่วงสำหรับประมาณค่าเฉลี่ยประชากร ภายใต้การสุ่มตัวอย่าง แบบใช้ความน่าจะเป็นในการเลือกไม่เท่ากัน กรณีที่สัดส่วนตัวอย่างมีขนาดเล็ก ผู้ช่วยศาสตราจารย์ ชูเกียรติ โพนแก้ว	10
การศึกษาคุณลักษณะของพอลิเมอร์ผสม อัลจินต – พอลิไวนิลแอลกอฮอล์ เพื่อประยุกต์ใช้ทางชีวการแพทย์ ดร.เกษร เมรัตน์	19
ปริมาณการสั่งซื้อเปลือกหอยทะเลที่เหมาะสม กรณีศึกษา ห้างหุ้นส่วนจำกัด พัฒนยนต์อโต้เซลล์ สรธร ศิริสวัสดิ์	29
การเตรียมตัวเร่งปฏิกิริยาจากโพแทสเซียมคาร์บอเนตบนไฮดรอกซีอะพาไทต์สำหรับการผลิต น้ำมันไบโอดีเซลจากน้ำมันปาล์ม อัญชนา พาลสิงห์	37
การลดความสูญเสียเปล่ากระบวนการผลิตอิฐบล็อก กรณีศึกษา: ร้านคำพุดก่อสร้าง ทานตะวัน เพ็ญมุล	44
การประมาณค่าสภาพต้านทานไฟฟ้าของดินจากข้อมูลทางอุตุนิยมวิทยา ผู้ช่วยศาสตราจารย์ สุวิทย์ เพชรห้วยลึก	54
การอบแห้งผลหมากรากไม้ด้วยตู้อบแห้งพลังงานแสงอาทิตย์ร่วมกับพลังงานชีวมวล ธีรเดช ใหญ่บ่ก	65
การเตรียมตัวเร่งปฏิกิริยาโซเดียมซิลิเกตจากเถ้าขานอ้อยสำหรับการผลิตไบโอดีเซล อรรคพล แสงพิพิธ	70

การประมาณค่าแบบช่วงสำหรับประมาณค่าเฉลี่ยประชากร ภายใต้การสุ่มตัวอย่างแบบใช้ความน่าจะเป็นในการเลือกไม่เท่ากัน กรณีที่สัดส่วนตัวอย่างมีขนาดเล็ก

Confidence Interval for estimating Population Mean under Unequal Probability Sampling with Negligible Sampling Fraction

ชูเกียรติ โพนแก้ว*

บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อศึกษาและเปรียบเทียบช่วงความเชื่อมั่นที่เป็นเชิงเส้น และช่วงความเชื่อมั่นที่ไม่เป็นเชิงเส้นสำหรับประมาณค่าเฉลี่ยประชากร ช่วงความเชื่อมั่นที่นำเสนอศึกษาภายใต้การสุ่มตัวอย่างแบบใช้ความน่าจะเป็นในการเลือกไม่เท่ากันและไม่ใส่คืน รวมถึงสัดส่วนตัวอย่างมีขนาดเล็ก ผู้วิจัยเปรียบเทียบประสิทธิภาพของช่วงความเชื่อมั่นที่นำเสนอ โดยใช้เทคนิคมอนติคาร์โลซิมูเลชัน และใช้ขนาดตัวอย่างทั้งหมด 5 ระดับ ประกอบด้วย 30, 35, 40, 45, และ 50 เกณฑ์ที่ใช้ในการพิจารณาประสิทธิภาพช่วงความเชื่อมั่นที่นำเสนอได้แก่ ความกว้างเฉลี่ยของช่วงความเชื่อมั่นที่สัมพันธ์กับความเชื่อมั่น เท่ากับ 0.95 ผลการเปรียบเทียบพบว่าทุกระดับขนาดตัวอย่างช่วงความเชื่อมั่นที่ไม่เป็นเชิงเส้น มีประสิทธิภาพดีกว่าช่วงความเชื่อมั่นที่เป็นเชิงเส้น

ABSTRACT

This research aim to proposes and compare the linear confidence interval and nonlinear confidence interval for estimating population mean. The propose confidence interval are investigate under unequal probability sampling without replacement with overall negligible sampling fraction. The author compares the efficiency of confidence interval by using Monte carol simulation and select sample with 5 levels of sample size there are 30, 35, 40, 45 and 50. The considered criterions are based on average length of confidence interval with 0.95 confidence coefficient. The result shows that for all level of sample size the nonlinear confidence interval is more efficiency than linear confidence interval.

คำสำคัญ: ช่วงความเชื่อมั่น, สัดส่วนตัวอย่าง, สัมประสิทธิ์ความเชื่อมั่น

Keywords: Confidence interval, Sampling fraction, Confidence coefficient

บทนำ

การประมาณค่า เป็นวิธีการทางสถิติที่ใช้ในการประมาณพารามิเตอร์ (Parameter) ที่สนใจศึกษาของประชากร โดยใช้ตัวสถิติ (Statistic) ที่คำนวณได้จากตัวอย่าง การประมาณค่าแบ่งออกเป็น การประมาณค่าแบบจุด และการประมาณค่าแบบช่วง สำหรับการประมาณค่าแบบจุดเป็นวิธีการประมาณค่าพารามิเตอร์ที่สนใจศึกษาจากค่าสถิติเพียงค่าเดียว ซึ่งค่าที่ได้ อาจมีค่าเท่ากับ หรือแตกต่างจากค่าจริงของพารามิเตอร์ ส่วนการประมาณค่าแบบช่วง เป็นการใช้อัตราตัวอย่างสร้างตัวสถิติที่จะครอบคลุมพารามิเตอร์ขึ้นมาสองค่า และเรียกตัวสถิติดังกล่าวว่าขีดความเชื่อมั่นล่าง และขีดความเชื่อมั่นบน โดยการประมาณค่าแบบช่วงมีขั้นตอนในการสร้างดังนี้ กำหนดให้ θ แทนพารามิเตอร์ที่สนใจศึกษา $\hat{\theta}$ แทนตัวประมาณค่าแบบจุดของ θ และ $E(\hat{\theta}) = \theta$ กำหนด $\hat{V}(\hat{\theta})$ แทนตัวประมาณค่าความแปรปรวนของ $\hat{\theta}$ นอกจากนี้กำหนดให้ Z (Z- statistic) เป็นตัวแปรสุ่ม และมีค่าตามที่แสดงในสมการที่ (1)

$$Z = \frac{\hat{\theta} - E(\hat{\theta})}{\sqrt{\hat{V}(\hat{\theta})}} = \frac{\hat{\theta} - \theta}{\sqrt{\hat{V}(\hat{\theta})}} \quad (1)$$

*ผู้ช่วยศาสตราจารย์ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏเพชรบูรณ์ อีเมล : chugiat.pon@pcru.ac.th

กรณีที่ขนาดตัวอย่างมีขนาดใหญ่ ($n \geq 30$) และใช้ทฤษฎีเข้าสู่ส่วนกลาง (Central limit theorem) สามารถแสดงได้ว่า $Z \sim \text{app}N(0,1)$ และเมื่อหาช่วงความเชื่อมั่นตามวิธีของ Casella & Berger (2002) จะได้ว่าช่วงความเชื่อมั่นของ θ ที่ระดับ $(1-\alpha)100\%$ เขียนแทนด้วยสัญลักษณ์ $C.I.(\theta)$ มีค่าตามที่แสดงในสมการที่ (2)

$$C.I.(\theta) = \left[\hat{\theta} - z_{1-\frac{\alpha}{2}} \sqrt{\hat{V}(\hat{\theta})}, \hat{\theta} + z_{1-\frac{\alpha}{2}} \sqrt{\hat{V}(\hat{\theta})} \right] \quad (2)$$

ค่าเฉลี่ยประชากร (Population mean) เป็นหนึ่งในพารามิเตอร์ที่นักสถิติ หรือนักวิจัยสนใจศึกษา และมีค่าเท่ากับผลรวมของคุณลักษณะที่สนใจศึกษาทุกหน่วยในประชากรหารด้วยขนาดของประชากร ดังนั้นกรณีที่ประชากรมีขนาดใหญ่ หรือไม่ทราบขนาดที่แท้จริงของประชากร ส่งผลให้ไม่สามารถคำนวณค่าเฉลี่ยประชากรได้ พิจารณาประชากร $U=\{1,2,\dots,N\}$ ขนาด N กำหนดให้ y แทนตัวแปรที่สนใจศึกษา และพารามิเตอร์ที่ต้องการทราบค่าได้แก่ค่าเฉลี่ยประชากรเขียนแทนด้วยสัญลักษณ์ $\bar{Y} = N^{-1} \sum_{i \in U} y_i$ เมื่อ y_i แทนค่าสังเกตของ

หน่วยตัวอย่างที่ i เมื่อ $i \in U$ สมมติว่าตัวอย่าง s ขนาด n ถูกสุ่มจากประชากร โดยใช้แผนการสุ่มตัวอย่างแบบใช้ความน่าจะเป็นในการเลือกไม่เท่ากันและใส่คืน Hansen & Horvitz (1946) ได้นำเสนอตัวประมาณค่าเฉลี่ยประชากร รวมถึงตัวประมาณค่าความแปรปรวนของตัวประมาณค่าที่นำเสนอครั้งนี้ กำหนดให้ $\hat{Y}_{HH}$ แทนตัวประมาณค่าของ Hansen & Horvitz (1946) และกำหนดค่าตามที่แสดงในสมการที่ (3)

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

เมื่อ $b_i = f(y_i)$ ส่วนตัวประมาณค่าความแปรปรวนของ $\hat{Y}_{HH}$ หรือ $\hat{V}(\hat{Y}_{HH})$ มีค่าตามที่แสดงในสมการที่ (4)

$$\hat{V}(\hat{Y}_{HH}) = \frac{1}{N^2} \left(\frac{n}{n-1} \right) \left[\sum_{i \in s} \hat{b}_i^2 - \frac{1}{n} \left(\sum_{i \in s} \hat{b}_i \right)^2 \right] \quad (4)$$

เมื่อ $\hat{b}_i$ แทนตัวประมาณค่าของ b_i ถ้า b_i มีพารามิเตอร์ที่ไม่ทราบค่า และ $\hat{b}_i = b_i$ ในกรณีอื่นๆ ต่อมา Horvitz & Thompson (1952) ได้ขยายตัวประมาณค่าของ Hansen & Horvitz (1946) ไปยังแผนการสุ่มตัวอย่างแบบใช้ความน่าจะเป็นในการเลือกไม่เท่ากันและไม่ใส่คืน ซึ่งตัวประมาณค่าที่ Horvitz & Thompson (1952) นำเสนออยู่ในรูปตัวประมาณค่าแบบเชิงเส้น เขียนแทนด้วย $\hat{Y}_{HT}$ และมีค่าตามที่แสดงในสมการที่ (5)

$$\hat{Y}_{HT} = \frac{1}{N} \sum_{i \in s} \frac{y_i}{\pi_i} \quad (5)$$

เมื่อ $\pi_i = P(i \in s)$ และตัวประมาณค่าความแปรปรวนของ $\hat{Y}_{HT}$ มีค่าตามที่แสดงในสมการที่ (6)

$$\hat{V}(\hat{Y}_{HT}) = \frac{1}{N^2} \left[\sum_{i \in s} \frac{1-\pi_i}{\pi_i^2} y_i^2 + \sum_{i \in s} \sum_{j \in s, j \neq i} \frac{\pi_{ij} - \pi_i \pi_j}{\pi_{ij} \pi_i \pi_j} y_i y_j \right] \quad (6)$$

เมื่อ $\pi_{ij} = P(i, j \in s)$ นอกจากนี้เมื่ออาศัยวิธีการตามที่ได้กล่าวถึงในสมการที่ (1) และ (2) ช่วงความเชื่อมั่น $(1-\alpha)100\%$ ของ $\bar{Y}$ ที่สร้างจากตัวประมาณค่าของ Horvitz & Thompson (1952) มีค่าตามที่แสดงในสมการที่ (7)

$$C.I_{(1)}(\bar{Y}) = \left[\hat{\bar{Y}}_{HT} - z_{1-\frac{\alpha}{2}} \sqrt{\hat{V}(\hat{\bar{Y}}_{HT})}, \hat{\bar{Y}}_{HT} + z_{1-\frac{\alpha}{2}} \sqrt{\hat{V}(\hat{\bar{Y}}_{HT})} \right] \quad (7)$$

ต่อมา Hájek (1971) ได้นำเสนอตัวประมาณค่าแบบไม่เป็นเชิงเส้น เพื่อใช้ในการประมาณค่า $\bar{Y}$ และมีค่าตามที่แสดงในสมการที่ (8)

$$\hat{\bar{Y}}_{Ha'jek} = \sum_{i \in S} \frac{y_i}{\pi_i} / \sum_{i \in S} \frac{1}{\pi_i} \quad (8)$$

Särndal Swenson & Wretment (1992) กล่าวว่ากรณีที่ π_i มีความสัมพันธ์ในทิศทางตรงข้ามกับ y_i ตัวประมาณค่าของ Hájek (1971) จะมีประสิทธิภาพดีกว่าตัวประมาณค่าของ Horvitz & Thompson (1952) ต่อมา Tillé & Matei (2016) ได้ทำการเปรียบเทียบประสิทธิภาพของตัวประมาณค่าทั้งสองตัวภายใต้เงื่อนไขดังกล่าว อย่างไรก็ตาม $\hat{\bar{Y}}_{Ha'jek}$ อยู่ในรูปของตัวประมาณค่าแบบไม่เป็นเชิงเส้น ดังนั้นจำเป็นต้องใช้เทเลอร์แบบเชิงเส้นในการปรับตัวประมาณค่าดังกล่าวให้อยู่ในรูปแบบเชิงเส้น ก่อนการหาตัวประมาณค่าความแปรปรวน และมีค่าตามที่แสดงในสมการที่ (9)

$$\hat{V}(\hat{\bar{Y}}_{Ha'jek}) = \frac{1}{N^2} \left[\sum_{i \in S} \frac{1-\pi_i}{\pi_i^2} (y_i - \hat{\bar{Y}}_{Ha'jek})^2 + \sum_{i \in S} \sum_{j \in S, j \neq i} \frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j} (y_i - \hat{\bar{Y}}_{Ha'jek})(y_j - \hat{\bar{Y}}_{Ha'jek}) \right] \quad (9)$$

เมื่ออาศัยวิธีการตามที่ได้กล่าวถึงในสมการที่ (1) และ สมการที่ (2) ช่วงความเชื่อมั่นที่ระดับ $(1-\alpha)100\%$ ของ $\bar{Y}$ ที่สร้างจากตัวประมาณค่าของ Hájek (1971) มีค่าตามที่แสดงในสมการที่ (10)

$$C.I_{(2)}(\bar{Y}) = \left[\hat{\bar{Y}}_{Ha'jek} - z_{1-\frac{\alpha}{2}} \sqrt{\hat{V}(\hat{\bar{Y}}_{Ha'jek})}, \hat{\bar{Y}}_{Ha'jek} + z_{1-\frac{\alpha}{2}} \sqrt{\hat{V}(\hat{\bar{Y}}_{Ha'jek})} \right] \quad (10)$$

จากสมการที่ (7) และสมการที่ (10) จะเห็นว่าช่วงความเชื่อมั่นของ $\bar{Y}$ ที่สร้างจากตัวประมาณค่าของ Horvitz & Thompson (1952) และตัวประมาณค่าของ Hájek (1971) ค่อนข้างมีความยุ่งยากในการคำนวณ เนื่องจากตัวประมาณค่าความแปรปรวน $\hat{V}(\hat{\bar{Y}}_{HT})$ และ $\hat{V}(\hat{\bar{Y}}_{Ha'jek})$ ต้องการค่า π_{ij} ซึ่งปกติจะมีความซับซ้อนในการกำหนดค่าที่แท้จริง (Berger, 2004) อย่างไรก็ตามกรณีที่สัดส่วนตัวอย่าง ($f = nN^{-1}$) มีขนาดเล็ก (ขนาดตัวอย่างน้อยกว่าหรือเท่ากับ 5% ของขนาดประชากร) ตัวประมาณค่าความแปรปรวนของตัวประมาณค่าเฉลี่ยประชากรภายใต้แผนการสุ่มตัวอย่างแบบใช้ความน่าจะเป็นในการเลือกไม่เท่ากันและไม่ใส่คืน สามารถประมาณค่าด้วยตัวประมาณค่าที่ Hansen & Horvitz (1946) นำเสนอตามที่ได้แสดงในสมการที่ (4)

ดังนั้นการวิจัยในครั้งนี้ผู้วิจัยต้องการพัฒนาช่วงความเชื่อมั่นสำหรับประมาณค่าเฉลี่ยประชากรที่สร้างตัวประมาณค่าของ Horvitz & Thompson (1952) และตัวประมาณค่าของ Hájek (1971) ตามที่ได้แสดงในสมการที่ (7) และสมการที่ (10) โดยการประมาณค่าความแปรปรวน ของตัวประมาณที่จุดที่ Horvitz & Thompson (1952) และ Hájek (1971) ที่นำเสนอด้วยวิธีการของ Hansen & Horvitz (1946) และเรียกช่วงความเชื่อมั่นที่ได้ว่าช่วงความเชื่อมั่นที่เป็นเชิงเส้น และช่วงความเชื่อมั่นที่ไม่เป็นเชิงเส้นตามลำดับ

วัตถุประสงค์

การวิจัยในครั้งนี้มีวัตถุประสงค์เพื่อ ศึกษาและเปรียบเทียบประสิทธิภาพช่วงความเชื่อมั่นสำหรับประมาณค่าเฉลี่ยประชากรภายใต้การสุ่มตัวอย่างแบบใช้ความน่าจะเป็นในการเลือกไม่เท่ากันและไม่ใส่คืน

วิธีดำเนินการวิจัย

การนำเสนอ และเปรียบเทียบประสิทธิภาพช่วงความเชื่อมั่นสำหรับประมาณค่าเฉลี่ยประชากรมีขั้นตอนดังนี้

1. การนำเสนอช่วงความเชื่อมั่นสำหรับประมาณค่าเฉลี่ยประชากร

สำหรับช่วงความเชื่อมั่นที่ผู้วิจัยนำเสนอ ได้พัฒนาจากช่วงความเชื่อมั่นที่สร้างจากช่วงความเชื่อมั่นของ Horvitz & Thompson (1952) และช่วงความเชื่อมั่นของ Hájek (1971) โดยใช้วิธีการประมาณค่าความแปรปรวนของ Hansen & Horvitz (1946) ตามที่แสดงในสมการที่ (4) ประมาณค่าความแปรปรวนของ $\hat{Y}_{HT}$ และ $\hat{Y}_{Ha'jek}$ เขียนแทนด้วย สัญลักษณ์ $\hat{V}_1$ และ $\hat{V}_2$ ตามลำดับ จากนั้นหาช่วงความเชื่อมั่น โดยแทนค่า $\hat{V}(\hat{Y}_{HT})$ ในสมการที่ (7) ด้วย $\hat{V}_1$ และเรียกช่วงความเชื่อมั่นที่ได้ว่า ช่วงความเชื่อมั่นที่เป็นเชิงเส้น เขียนแทนด้วย สัญลักษณ์ $Adjusted\ C.I._{(1)}(\bar{Y})$ ต่อมาแทนค่า $\hat{V}(\hat{Y}_{Ha'jek})$ ในสมการที่ (10) ด้วย $\hat{V}_2$ และเรียกช่วงความเชื่อมั่นที่ได้ว่า ช่วงความเชื่อมั่นที่ไม่เป็นเชิงเส้น เขียนแทนด้วย สัญลักษณ์ $Adjusted\ C.I._{(2)}(\bar{Y})$

2. การเปรียบเทียบประสิทธิภาพช่วงความเชื่อมั่นที่นำเสนอโดยใช้เทคนิคมอนติคาร์โลซิมูลेशन

2.1 กำหนดข้อมูลประชากร

ผู้วิจัยพิจารณาข้อมูลประชากร กรณีที่ y_i มีความสัมพันธ์ในทิศทางตรงข้ามกับ π_i เมื่อ $i \in U$ โดยกำหนดข้อมูลประชากรตามการศึกษาของ Tillé & Matei (2016) ดังนี้ กำหนด $N=1,000$ ค่า $k_i=i$ เมื่อ $i=1,2,...,N$ จากนั้นกำหนดค่า π_i โดยใช้ฟังก์ชัน inclusion probabilities () ใน package “Sampling” ของโปรแกรม R โดยกำหนด $\pi = inclusionprobabilities(k,n)$ เมื่อ $\pi = [\pi_1, \pi_2, ..., \pi_N]'$ ค่า $k = [k_1, k_2, ..., k_N]'$ สุดท้ายกำหนดค่า y_i โดยที่ $y_i = \pi_i^{-1}$ เมื่อ $i=1,2,...,N$

2.2 คำนวณขีดความเชื่อมั่นล่าง และขีดความเชื่อมั่นบนที่สัมพันธ์กับความเชื่อมั่นเท่ากับ 0.95

เทคนิคมอนติคาร์โลซิมูลेशनกำหนดการทำซ้ำทั้งหมด 50,000 หรือ $M=50,000$

สำหรับ ค่า $m=1$ ถึง M

– สุ่มตัวอย่าง s ขนาด $n=30$

– คำนวณค่า $L_k[m]$ และ $U_k[m]$, $k=1,2$

เมื่อ $L_k[m]$ และ $U_k[m]$ แทน ขีดความเชื่อมั่นล่าง และขีดความเชื่อมั่นบนของช่วงความเชื่อมั่นในรอบที่ m

2.3 คำนวณค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่น

กำหนดให้ AL_1 และ AL_2 แทนความกว้างเฉลี่ยของช่วงความเชื่อมั่นที่เป็นเชิงเส้น และช่วงความเชื่อมั่นที่ไม่เป็นเชิงเส้นตามลำดับ และคำนวณค่าจากสูตร

$$AL_k = \frac{1}{M} \sum_{m=1}^M (U_k[m] - L_k[m]), \quad k=1,2 \quad (11)$$

จากนั้นทำซ้ำขั้นตอนที่ 2.1 ถึง 2.3 แต่เปลี่ยนขนาดตัวอย่างจาก $n=30$ เป็น 35, 40, 45 และ 50 ตามลำดับ

2.4 เกณฑ์การเปรียบเทียบประสิทธิภาพช่วงความเชื่อมั่น

สำหรับเกณฑ์การเปรียบเทียบประสิทธิภาพของช่วงความเชื่อมั่นที่นำเสนอ ผู้วิจัยใช้เกณฑ์ตามที่กำหนดตามบทนิยามที่ 1

บทนิยาม 1 กำหนดให้ $C.I._A(\bar{Y})$ และ $C.I._B(\bar{Y})$ แทนช่วงความเชื่อมั่นสำหรับประมาณค่าเฉลี่ยประชากร

จะกล่าวว่า $C.I._A(\bar{Y})$ มีประสิทธิภาพดีกว่า $C.I._B(\bar{Y})$ ก็ต่อเมื่อ $AL_A < AL_B$ เมื่อ AL_A และ AL_B แทนความกว้างเฉลี่ยของ $C.I._A(\bar{Y})$ และ $C.I._B(\bar{Y})$ ตามลำดับ

ผลการศึกษา

การวิจัยในครั้งนี้มีวัตถุประสงค์เพื่อ นำเสนอช่วงความเชื่อมั่นสำหรับการประมาณค่าเฉลี่ยประชากร โดยไม่จำเป็นต้องทราบค่า π_j เนื่องจากค่าดังกล่าวค่อนข้างมีความยุ่งยากในการคำนวณค่าภายใต้แผนการสุ่มแบบใช้ความน่าจะเป็นในการเลือกไม่เท่ากันและไม่ใส่คืน ผลการศึกษาเป็นดังนี้

1. ช่วงความเชื่อมั่นที่น่าเสนอ

1.1 ช่วงความเชื่อมั่นที่เป็นเชิงเส้น

จากสมการ (5) สามารถเขียนตัวประมาณค่าของ Horvitz & Thompson (1952) ให้อยู่ในรูป

$$\hat{Y}_{HT} = \hat{Y}_1 = \frac{1}{N} \sum_{i \in s} \frac{y_i}{\pi_i} = \frac{1}{N} \sum_{i \in s} b_i \quad (12)$$

เมื่อ $b_i = f(y_i) = y_i \pi_i^{-1}$ จากนั้นหาตัวประมาณค่าความแปรปรวนของ $\hat{Y}_1$ ด้วยวิธีของ Hansen & Horvitz (1946) และเขียนแทนด้วยสัญลักษณ์ $\hat{V}_1$ เนื่องจากค่า $b_i = y_i \pi_i^{-1}$ ไม่ปรากฏตัวพารามิเตอร์ที่ไม่ทราบค่า ดังนั้น $\hat{b}_i = y_i \pi_i^{-1}$ และแทนค่าในสมการที่ (4) ได้ว่า

$$\begin{aligned} \hat{V}_1 &= \frac{1}{N^2} \left(\frac{n}{n-1} \right) \left[\sum_{i \in s} \left(\frac{y_i}{\pi_i} \right)^2 - \frac{1}{n} \left(\sum_{i \in s} \frac{y_i}{\pi_i} \right)^2 \right] \\ &= \frac{1}{N^2} \left(\frac{n}{n-1} \right) \left[\sum_{i \in s} \tilde{y}_{1i}^2 - \frac{1}{n} \left(\sum_{i \in s} \tilde{y}_{1i} \right)^2 \right] \end{aligned} \quad (13)$$

เมื่อ $\tilde{y}_{1i} = y_i \pi_i^{-1}$ จากสมการที่ (12) และ (13) กำหนดตัวแปรสุ่ม $Z = (\hat{Y}_1 - \bar{Y}) / \sqrt{\hat{V}_1}$ กรณีที่ตัวอย่างมีขนาดใหญ่ และใช้ทฤษฎีเข้าสู่ส่วนกลาง สามารถแสดงได้ว่า

$$Z = \frac{(\hat{Y}_1 - \bar{Y})}{\sqrt{\hat{V}_1}} \sim \text{app}N(0,1) \quad (14)$$

ดังนั้นช่วงความเชื่อมั่นที่เป็นเชิงเส้น เพื่อใช้ประมาณค่าเฉลี่ยประชากร โดยใช้วิธีของ Casella & Berger (2002) มีค่าตามที่แสดงในทฤษฎี 1

ทฤษฎี 1 กำหนดให้ y แทนตัวแปรที่สนใจศึกษา และ $\bar{Y}$ แทนค่าเฉลี่ยประชากร สมมติว่า ตัวอย่าง s ขนาด n ถูกสุ่มจากประชากร U ขนาด N โดยแผนการสุ่มตัวอย่างแบบใช้ความน่าจะเป็นในการเลือกไม่เท่ากันและไม่ใส่คืน และ n มีขนาดใหญ่รวมถึงไม่มีการสูญหายของข้อมูล กำหนดค่า $\hat{Y}_1$ ตามที่แสดงในสมการที่ (12) และ $\hat{V}_1$ ตามที่แสดงในสมการที่ (13) ช่วงความเชื่อมั่นที่เป็นเชิงเส้นเพื่อใช้ประมาณค่า $\bar{Y}$ ที่ระดับ $(1-\alpha)100\%$ เขียนแทนด้วย $\text{Adjusted } C.I._{(1)}(\bar{Y})$ มีค่าตามที่แสดงในสมการที่ (15)

$$\text{Adjusted } C.I._{(1)}(\bar{Y}) = \left[\hat{Y}_1 - z_{1-\frac{\alpha}{2}} \sqrt{\hat{V}_1}, \hat{Y}_1 + z_{1-\frac{\alpha}{2}} \sqrt{\hat{V}_1} \right] = [L_1, U_1] \quad (15)$$

$$\text{เมื่อ } L_1 = \hat{Y}_1 - z_{1-\frac{\alpha}{2}} \sqrt{\hat{V}_1} \text{ และ } U_1 = \hat{Y}_1 + z_{1-\frac{\alpha}{2}} \sqrt{\hat{V}_1}$$

พิสูจน์ กำหนดค่า $\hat{Y}_1$ และ $\hat{V}_1$ ตามที่แสดงในสมการที่ (12) และ (13) ตามลำดับ จากสมการที่ (14)

กรณีที่ n มีขนาดใหญ่ $Z = \frac{(\hat{Y}_1 - \bar{Y})}{\sqrt{\hat{V}_1}} \sim \text{app}N(0,1)$ เมื่อใช้วิธีของ Casella & Berger (2002) ได้ว่า

$$P(z_{\frac{\alpha}{2}} < Z < z_{1-\frac{\alpha}{2}}) = 1 - \alpha$$

$$P(z_{\frac{\alpha}{2}} < \frac{\hat{Y}_1 - \bar{Y}}{\sqrt{\hat{V}(\hat{Y}_1)}} < z_{1-\frac{\alpha}{2}}) = 1 - \alpha$$

$$P(-z_{1-\frac{\alpha}{2}} < \frac{\hat{Y}_1 - \bar{Y}}{\sqrt{\hat{V}(\hat{Y}_1)}} < z_{1-\frac{\alpha}{2}}) = 1 - \alpha$$

ดังนั้นช่วงความเชื่อมั่นที่เป็นเชิงเส้น เพื่อใช้ประมาณค่า $\bar{Y}$ ที่ระดับ $(1-\alpha)100\%$ มีค่าเป็น

$$\text{Adjusted C.I.}_{(1)}(\bar{Y}) = \left[\hat{Y}_1 - z_{1-\frac{\alpha}{2}} \sqrt{\hat{V}_1}, \hat{Y}_1 + z_{1-\frac{\alpha}{2}} \sqrt{\hat{V}_1} \right] = [L_1, U_1]$$

$$\text{เมื่อ } L_1 = \hat{Y}_1 - z_{1-\frac{\alpha}{2}} \sqrt{\hat{V}_1} \text{ และ } U_1 = \hat{Y}_1 + z_{1-\frac{\alpha}{2}} \sqrt{\hat{V}_1}$$

1.2 ช่วงความเชื่อมั่นที่ไม่เป็นเชิงเส้น

หัวข้อที่แล้วผู้วิจัยได้นำเสนอช่วงความเชื่อมั่นที่เป็นเชิงเส้น เพื่อใช้ในการประมาณ $\bar{Y}$ ต่อไปผู้วิจัยจะได้กล่าวถึงช่วงความเชื่อมั่นที่ไม่เป็นเชิงเส้นดังนี้ จากสมการที่ (8) Hájek (1971) ได้นำเสนอตัวประมาณค่าแบบไม่เป็นเชิงเส้น เพื่อใช้ประมาณค่า $\bar{Y}$ ตามที่แสดงในสมการที่ (16)

$$\hat{Y}_{Ha'jek} = \sum_{i \in s} \frac{y_i}{\pi_i} / \sum_{i \in s} \frac{1}{\pi_i} \quad (16)$$

เนื่องจาก $\hat{Y}_{Ha'jek}$ เป็นตัวประมาณค่าแบบไม่เป็นเชิงเส้น ดังนั้นการหาตัวประมาณค่าความแปรปรวน จำเป็นต้องใช้เทเลอร์แบบเชิงเส้นในการปรับตัวประมาณค่าดังกล่าวให้เป็นเชิงเส้นก่อนหาตัวประมาณค่าความแปรปรวน และเมื่อใช้วิธีเทเลอร์แบบเชิงเส้น สามารถเขียน $\hat{Y}_{Ha'jek}$ ให้อยู่ในรูป

$$\hat{Y}_{Ha'jek} = \hat{Y}_2 \approx \frac{1}{N} \sum_{i \in s} \frac{1}{\pi_i} (y_i - \bar{Y}) = \frac{1}{N} \sum_{i \in s} b_i \quad (17)$$

เมื่อ $b_i = \frac{1}{\pi_i} (y_i - \bar{Y})$ จากนั้นหาตัวประมาณค่าความแปรปรวนของ $\hat{Y}_{Ha'jek}$ ด้วยวิธีของ Hansen & Horvitz (1946) และ

เขียนแทนด้วยสัญลักษณ์ $\hat{V}_2$ อย่างไรก็ตาม ค่า $b_i = \frac{1}{\pi_i} (y_i - \bar{Y})$ มีพารามิเตอร์ที่ไม่ทราบค่าคือ $\bar{Y}$ และสามารถประมาณค่า

ด้วย $\hat{Y}_{Ha'jek}$ ดังนั้น $\hat{b}_i = \frac{1}{\pi_i} (y_i - \hat{Y}_{Ha'jek})$ และแทนค่า $\hat{b}_i$ ในสมการที่ (4) ได้ว่า

$$\hat{V}_2 = \frac{1}{N^2} \left(\frac{n}{n-1} \right) \left[\sum_{i \in s} \left(\frac{1}{\pi_i} (y_i - \hat{Y}_{Ha'jek}) \right)^2 - \frac{1}{n} \left(\sum_{i \in s} \frac{1}{\pi_i} (y_i - \hat{Y}_{Ha'jek}) \right)^2 \right]$$

$$= \frac{1}{N^2} \left(\frac{n}{n-1} \right) \left[\sum_{i \in S} \tilde{y}_{2i}^2 - \frac{1}{n} \left(\sum_{i \in S} \tilde{y}_{2i} \right)^2 \right] \quad (18)$$

เมื่อ $\tilde{y}_{2i} = \pi_i^{-1} (y_i - \hat{\gamma}_{Ha'jek})$ จากนั้นกำหนดตัวแปรสุ่ม $Z = (\hat{\gamma}_2 - \bar{Y}) / \sqrt{\hat{V}_2}$ กรณีที่ตัวอย่างมีขนาดใหญ่ และใช้ทฤษฎีเข้าสู่ส่วนกลาง สามารถแสดงได้ว่า

$$Z = \frac{(\hat{\gamma}_2 - \bar{Y})}{\sqrt{\hat{V}_2}} \sim \text{app}N(0,1) \quad (19)$$

ดังนั้นช่วงความเชื่อมั่นที่ไม่เป็นเชิงเส้น เพื่อใช้ประมาณค่าเฉลี่ยประชากร โดยใช้วิธีของ Casella & Berger (2002) มีค่าตามที่แสดงในทฤษฎี 2

ทฤษฎี 2 กำหนดให้ y แทนตัวแปรที่สนใจศึกษา และ $\bar{Y}$ แทนค่าเฉลี่ยประชากร สมมติว่า ตัวอย่าง s ขนาด n ถูกสุ่มจากประชากร U ขนาด N โดยแผนการสุ่มตัวอย่างแบบใช้ความน่าจะเป็นในการเลือกไม่เท่ากันและไม่ใส่คืน และ n มีขนาดใหญ่รวมถึงไม่มีการสูญหายของข้อมูล กำหนดค่า $\hat{\gamma}_2$ ตามที่แสดงในสมการที่ (16) และ $\hat{V}_2$ ตามที่แสดงในสมการที่ (18) ช่วงความเชื่อมั่นที่ไม่เป็นเชิงเส้นเพื่อใช้ประมาณค่า $\bar{Y}$ ที่ระดับ $(1-\alpha)100\%$ เขียนแทนด้วย *Adjusted C.I.*₍₂₎($\bar{Y}$) และกำหนดค่าตามที่แสดงในสมการที่ (20)

$$\text{Adjusted C.I.}_{(2)}(\bar{Y}) = \left[\hat{\gamma}_2 - z_{1-\frac{\alpha}{2}} \sqrt{\hat{V}_2}, \hat{\gamma}_2 + z_{1-\frac{\alpha}{2}} \sqrt{\hat{V}_2} \right] = [L_2, U_2] \quad (20)$$

$$\text{เมื่อ } L_2 = \hat{\gamma}_2 - z_{1-\frac{\alpha}{2}} \sqrt{\hat{V}_2} \text{ และ } U_2 = \hat{\gamma}_2 + z_{1-\frac{\alpha}{2}} \sqrt{\hat{V}_2}$$

การพิสูจน์ เช่นเดียวกับทฤษฎี 1 จากสมการที่ (15) และสมการที่ (20) สามารถเขียนสูตรทั่วไปของช่วงความเชื่อมั่นที่เป็นเชิงเส้น และช่วงความเชื่อมั่นที่ไม่เป็นเชิงเส้น ได้ดังนี้

$$\text{Adjusted C.I.}_{(k)}(\bar{Y}) = \left[\hat{\gamma}_k - z_{1-\frac{\alpha}{2}} \sqrt{\hat{V}_k}, \hat{\gamma}_k + z_{1-\frac{\alpha}{2}} \sqrt{\hat{V}_k} \right], k=1,2 \quad (21)$$

เมื่อ $\hat{\gamma}_1 = \hat{\gamma}_{HT}$ และ $\hat{\gamma}_2 = \hat{\gamma}_{Ha'jek}$ ส่วนค่า $\hat{V}_1$ และ $\hat{V}_2$ กำหนดค่าตามที่แสดงในสมการ (13) และ (18) ตามลำดับ

2. ผลการเปรียบเทียบประสิทธิภาพ

หัวข้อนี้ผู้วิจัยจะได้นำผลเสนอผลการเปรียบเทียบประสิทธิภาพของช่วงความเชื่อมั่นที่เป็นเชิงเส้น และช่วงความเชื่อมั่นที่ไม่เป็นเชิงเส้น โดยพิจารณาจากค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่น หรือค่า AL_k เมื่อ $k=1,2$ ตามที่กล่าวถึงในบทนิยามที่ 1 ส่วนผลการคำนวณค่า AL_k แสดงดังตารางที่ 1 ซึ่งผลจากตารางที่ 1 พบว่าในแต่ละระดับขนาดตัวอย่าง ค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่นที่ไม่เป็นเชิงเส้นมีค่าน้อยกว่าช่วงความเชื่อมั่นที่เป็นเชิงเส้น ดังนั้นโดยอาศัยบทนิยาม 1 จะได้ว่าช่วงความเชื่อมั่นที่ไม่เป็นเชิงเส้นมีประสิทธิภาพดีกว่าช่วงความเชื่อมั่นที่เป็นเชิงเส้น

ตารางที่ 1 ค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่นที่เป็นเชิงเส้น และไม่เป็นเชิงเส้น

n	AL_1	AL_2
30	234.95	125.62
35	228.76	109.31
40	164.59	87.12
45	150.27	82.72
50	135.63	73.77

สรุปผลและอภิปรายผล

การวิจัยในครั้งนี้มีวัตถุประสงค์ เพื่อนำเสนอ และเปรียบเทียบช่วงความเชื่อมั่นสำหรับประมาณค่าเฉลี่ยประชากร ภายใต้แผนการสุ่มตัวอย่างแบบใช้ความน่าจะเป็นในการเลือกไม่เท่ากันและไม่ใส่คืน โดยไม่จำเป็นต้องทราบค่า π_{ij} เนื่องจากค่าดังกล่าวค่อนข้างยุ่งยากในการกำหนดค่าภายใต้แผนการสุ่มดังกล่าว (Berger, 2004) โดยปกติภายใต้แผนการสุ่มตัวอย่างแบบใช้ความน่าจะเป็นในการเลือกไม่เท่ากันและไม่ใส่คืน ค่าเฉลี่ยประชากรสามารถประมาณค่าได้จากช่วงความเชื่อมั่นที่สร้างจากตัวประมาณค่าแบบจุด และตัวประมาณค่าความแปรปรวนของ Horvitz & Thompson (1952) และ Hájek (1971) ตามที่แสดงในสมการที่ (7) และสมการที่ (10) อย่างไรก็ตามตัวประมาณค่าความแปรปรวนที่ใช้สร้างช่วงความเชื่อมั่นดังกล่าวอยู่ในรูปฟังก์ชันของ π_{ij} ซึ่งปกติค่อนข้างมีความยุ่งยากในการคำนวณค่า ดังนั้นผู้วิจัยได้นำเสนอช่วงความเชื่อมั่นที่เป็นเชิงเส้น และช่วงความเชื่อมั่นที่ไม่เป็นเชิงเส้นที่พัฒนาจากช่วงความเชื่อมั่นที่กล่าวถึงในสมการที่ (7) และสมการที่ (10) ตามขั้นตอนต่อไปนี้ ขั้นแรกหาตัวประมาณค่าความแปรปรวน ของตัวประมาณค่าแบบจุดที่ Horvitz & Thompson (1952) และ Hájek (1971) นำเสนอในกรณีที่ดีที่สุดส่วนตัวอย่างมีขนาดเล็กด้วยวิธีของ Hansen & Horvitz (1946) ขั้นที่สองสร้างตัวแปรสุ่มที่อยู่ในรูปของฟังก์ชันของตัวประมาณค่าแบบจุด และตัวประมาณค่าความแปรปรวนที่ได้จากขั้นตอนแรก จากนั้นใช้ทฤษฎีเข้าสู่ส่วนกลาง ปรับการแจกแจงของตัวแปรสุ่มดังกล่าวให้มีการแจกแจงใกล้เคียงแบบปกติ ขั้นสุดท้ายสร้างช่วงความเชื่อมั่นที่เป็นเชิงเส้น และช่วงความเชื่อมั่นที่ไม่เป็นเชิงเส้นตามวิธีของ Casella & Berger (2002) ช่วงความเชื่อมั่นที่ผู้วิจัยนำเสนอไม่จำเป็นต้องทราบค่า π_{ij} เนื่องจากตัวประมาณค่าความแปรปรวนที่ได้จากวิธีของ Hansen & Horvitz (1946) ไม่อยู่ในรูปฟังก์ชันของค่า π_{ij} ทำให้สะดวกและง่ายในการประมาณค่าเฉลี่ยประชากร

การเปรียบเทียบประสิทธิภาพของช่วงความเชื่อมั่นที่นำเสนอ ผู้วิจัยพิจารณาจากค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่น ซึ่งจิราภรณ์ สิงฆ (2560) กล่าวว่าช่วงความเชื่อมั่นที่ดีต้องมีค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่นแคบที่สุด ผลการเปรียบเทียบพบว่าช่วงความเชื่อมั่นที่ไม่เป็นเชิงเส้นมีประสิทธิภาพดีกว่าช่วงความเชื่อมั่นที่เป็นเชิงเส้น เนื่องจากค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่นที่ไม่เป็นเชิงเส้นแคบกว่า (น้อยกว่า) ช่วงความเชื่อมั่นที่เป็นเชิงเส้น นอกจากนี้จากสมการที่ (10) จะเห็นว่าค่าเฉลี่ยของความกว้างของช่วงความเชื่อมั่น จะมีค่าเท่ากับค่าเฉลี่ยของผลต่างระหว่างขีดความเชื่อมั่นบน และขีดความเชื่อมั่นล่าง ซึ่งค่าดังกล่าวมีค่าขึ้นอยู่กับค่าของตัวอย่างที่ถูกสุ่มจากประชากร สำหรับขั้นตอนการแสดงตัวอย่างเชิงตัวเลข ผู้วิจัยกำหนดข้อมูลประชากร กรณีที่ y_i มีความสัมพันธ์ในทิศทางตรงข้ามกับ π_i ซึ่ง Tillé & Matei (2016) ได้แสดงให้เห็นว่าภายใต้เงื่อนไขดังกล่าว ตัวประมาณแบบจุดที่นำเสนอโดย Hájek (1971) จะมีประสิทธิภาพดีกว่าตัวประมาณค่าแบบจุดของ Horvitz & Thompson (1952) ซึ่งสอดคล้องกับผลการศึกษาของผู้วิจัย ที่พบว่าช่วงความเชื่อมั่นที่พัฒนาจากตัวประมาณค่าของ Hájek (1971) มีประสิทธิภาพดีกว่าช่วงความเชื่อมั่นที่พัฒนาจากตัวประมาณค่าของ Horvitz & Thompson (1952)

ข้อเสนอแนะ

1. ข้อเสนอแนะในการนำไปใช้

ช่วงความเชื่อมั่นที่ผู้วิจัยนำเสนอเหมาะสำหรับข้อมูลที่ทราบข้อมูลของตัวแปรช่วยก่อนเริ่มสุ่มตัวอย่าง เช่น การประมาณค่าผลผลิตเฉลี่ยของข้าวโพดเลี้ยงสัตว์ของเกษตรกรในเขตอำเภอหล่มสัก จังหวัดเพชรบูรณ์ ซึ่งผู้วิจัยอาจใช้ข้อมูลพื้นที่ที่ใช้ในการเพาะปลูกของเกษตรกรในเขตอำเภอหล่มสัก เป็นข้อมูลตัวแปรช่วยซึ่งข้อมูลดังกล่าว สามารถขอได้จากสำนักงานการเกษตรอำเภอหล่มสัก หรือสำนักงานสถิติจังหวัดเพชรบูรณ์ เป็นต้น

2. ข้อเสนอแนะในการทำวิจัยครั้งต่อไป

ผู้สนใจศึกษาอาจขยายวิธีการของผู้วิจัยไปยังกรณีทั่วไป และใช้ วิธีแจ๊คไนฟ์ (Jackknife method) หรือวิธีบูตสเตรป (วิธีบูตสเตรป) ในการประมาณค่าความแปรปรวน เป็นต้น

เอกสารอ้างอิง

- จิราภรณ์ สิงฆ. (2560). การประมาณค่าช่วงความเชื่อมั่นของพารามิเตอร์ของการแจกแจงค่าสุดขีดเมื่อข้อมูลถูกเซ็นเซอร์ประเภทที่ 1. วิทยานิพนธ์ วิทยาศาสตร์มหาบัณฑิต (สถิติประยุกต์) สาขาวิชาคณิตศาสตร์และสถิติ. คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์.
- Casella, G., and Berger, R.L. (2002). *Statistical inference*. Duxbury Press, second edition.
- Berger, Y. G. (2004). *A simple variance estimator for unequal probability sampling without replacement. Journal of Applied Statistics*. Vol.31 : 305-315.
- Hajek, J. (1971) *Comment on An essay on the logical foundations of survey sampling by Basu, D. in Foundations of Statistical Inference (Godambe, V.P. and Sprott, D.A. eds.)*, p. 236. Holt, Rinehart and Winston.
- Hansen, M.H. and Hurwitz, W.N. (1946). The problem of non-response in sample surveys. *Journal of the American Statistical Association*, 41, 517-529.
- Horvitz, D.G., and D.J. Thompson. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47, 663-685.
- Särndal, CE., Swensson, B. and Wretman, J. (1992). *Model assisted survey sampling*. New York : Springer-Verlag.
- Tillé, Y., Matei, M. (2016). *Sampling: Survey Sampling*. R package version 2.8. Accessed August 28, 2018. (2016), <https://CRAN.R-project.org/package=sampling>.

ประวัติผู้วิจัย

ชื่อ-ชื่อสกุล : (ภาษาไทย) นายชูเกียรติ โพนแก้ว

(ภาษาอังกฤษ) Mr. Chugiat Ponkaew

ตำแหน่ง : ผู้ช่วยศาสตราจารย์ หลักสูตรสาขาวิชาคณิตศาสตร์
คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏเพชรบูรณ์

ไปรษณีย์อิเล็กทรอนิกส์ (e-mail) : chugiat.pon@pcru.ac.th

วัน เดือน ปีเกิด : 26 ตุลาคม 2520

สถานที่เกิด : บ้านโนนทราย บ้านเลขที่ 2 หมู่ที่ 8 ตำบลคูซอด
อำเภอเมืองศรีสะเกษ จังหวัดศรีสะเกษ 33000

หน่วยงานที่ติดต่อได้สะดวก : สาขาวิชาคณิตศาสตร์ คณะวิทยาศาสตร์และเทคโนโลยี
มหาวิทยาลัยราชภัฏเพชรบูรณ์

83 หมู่ 11 ถนน สระบุรี-หล่มสัก ตำบลสะเดียง อำเภอเมือง
เพชรบูรณ์ จังหวัดเพชรบูรณ์ รหัสไปรษณีย์ 67000

โทรศัพท์ : 0 - 5671 - 7100 ต่อ 1407

วุฒิการศึกษา : การศึกษาระดับปริญญาตรี หลักสูตรวิทยาศาสตรบัณฑิต สาขาวิชา
คณิตศาสตร์ มหาวิทยาลัยบูรพา เมื่อปีการศึกษา 2543
สำเร็จการศึกษาระดับปริญญาโท หลักสูตรวิทยาศาสตรมหาบัณฑิต
สาขาวิชาสถิติประยุกต์ มหาวิทยาลัยขอนแก่น ปีการศึกษา 2547
สำเร็จการศึกษาระดับปริญญาเอก หลักสูตรปรัชญาดุษฎีบัณฑิต
สาขาวิชาสถิติประยุกต์ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้า
พระนครเหนือ กรุงเทพมหานคร ปีการศึกษา 2561