

บทที่ 1

บทนำ

ความเป็นมาของปัญหา

การสำรวจด้วยตัวอย่าง เป็นการเก็บรวบรวมข้อมูลจากตัวอย่าง เพื่อนำมาวิเคราะห์ และสรุปผลเกี่ยวกับคุณลักษณะของประชากร ตัวอย่างที่ใช้ควรเป็นตัวอย่างสุ่ม และมีขนาดเหมาะสม เลือจากกรอบตัวอย่างที่สมบูรณ์ ซึ่งอาจเป็นกรอบรายชื่อ หรือกรอบพื้นที่ การสำรวจด้วยตัวอย่าง จะให้ประสิทธิภาพสูงต้องอาศัยการออกแบบแผนเลือกตัวอย่างที่ดี ใช้วิธีประมวลผลและการวิเคราะห์ข้อมูลที่เหมาะสม โดยลักษณะการสำรวจด้วยตัวอย่างที่ดีต้องมีการระบุประชากรที่ชัดเจน ก่อนการสำรวจ จากนั้นต้องหาหรือสร้างกรอบตัวอย่างที่ดีและทันสมัย เพื่อใช้ในการสุ่มตัวอย่าง และเลือกใช้แผนการสุ่มให้เหมาะสมกับคุณลักษณะของประชากรในกรอบตัวอย่าง เช่น ถ้าประชากรมีความคล้ายคลึงกัน แผนการสุ่มที่เหมาะสมก็คือแผนการสุ่มตัวอย่างแบบง่าย เป็นต้น นอกจากนี้ยังต้องคำนึงถึงข้อจำกัดต่าง ๆ สำหรับการวางแผนการสำรวจ เช่น เวลา ค่าใช้จ่าย บุคลากร การประมวลผล และการวิเคราะห์ข้อมูล เพื่อให้แผนการสำรวจอยู่ในวิสัยที่สามารถปฏิบัติได้จริง

การสำรวจด้วยตัวอย่างมีเป้าหมายอยู่ที่การหาตัวแทนหรือตัวอย่างที่ดีของประชากร เพื่อการอนุมานจากตัวอย่างที่ได้จากการสำรวจไปยังประชากรที่สนใจศึกษาหรือประชากรเป้าหมาย (Target Population) อาจจะเป็นการทดสอบสมมติฐาน หรือการประมาณค่า เช่น การประมาณค่าเฉลี่ย ผลรวม สัดส่วน เป็นต้น ผลลัพธ์ที่ได้จากการสำรวจด้วยตัวอย่างอาจมีความคลาดเคลื่อนจากค่าที่แท้จริงของประชากรเป้าหมายได้ และหากความคลาดเคลื่อนมีค่าสูงเกินไป การสำรวจก็เป็นสิ่งไร้ความหมาย หรือใช้ประโยชน์ต่อไปไม่ได้ไม่คุ้มค่าที่ควร

การสำรวจด้วยตัวอย่าง มักเกิดความคลาดเคลื่อนจากการสำรวจ ซึ่งความคลาดเคลื่อนสามารถจำแนกเป็น 2 ประเภท คือ 1) ความคลาดเคลื่อนจากการสุ่มตัวอย่าง (Sampling Errors) ซึ่งทำให้ผลการสำรวจไม่ตรงกับความเป็นจริง เนื่องจากใช้ข้อมูลที่รวบรวมจากบางส่วนของประชากรในการสรุปผลเกี่ยวกับคุณลักษณะของประชากร และ 2) ความคลาดเคลื่อนที่ไม่ได้เกิดจากการสุ่มตัวอย่าง (Nonsampling Errors) อาจเนื่องมาจากความคลาดเคลื่อนจากการประมวลผล ความคลาดเคลื่อนจากการวัด ความคลาดเคลื่อนคัมรวมหรือความคลาดเคลื่อนจากกรอบตัวอย่าง และความคลาดเคลื่อนจากการไม่ได้รับความร่วมมือ ซึ่งความคลาดเคลื่อนที่ไม่ได้เกิดจากการสุ่มตัวอย่างนั้นควบคุมได้ยาก มีรูปแบบไม่ชัดเจน และมีขนาดใหญ่ขึ้นเมื่อขนาด

ตัวอย่างเพิ่มขึ้น ดังนั้นหากไม่มีการควบคุมให้ตีความคลาดเคลื่อนประเภทนี้จะทำความเสียหายต่อการสำรวจได้มากกว่าความคลาดเคลื่อนจากการสุ่มตัวอย่าง โดยเฉพาะอย่างยิ่งในการสำรวจครัวเรือนที่มีขนาดใหญ่มา ก ๆ (ประชุม สุวดี, 2552, หน้า 494-500)

ความคลาดเคลื่อนที่ไม่ได้เกิดจากการสุ่มตัวอย่าง ที่เกิดจากการประมวลผล หรือการวัดสามารถลดได้ด้วยการอบรมพนักงานที่เกี่ยวข้องให้มีความเชี่ยวชาญในการเจนนับ การวัด การลงรหัสข้อมูล ตลอดจนมีการตรวจสอบความถูกต้องของข้อมูลก่อนการลงรหัสข้อมูล และการประมวลผล ส่วนความคลาดเคลื่อนที่เกิดจากการรอบตัวอย่างเป็นความคลาดเคลื่อนที่ควบคุม และหลีกเลี่ยงได้ยาก เนื่องจากการหาหรือสร้างกรอบตัวอย่างที่สมบูรณ์นั้นทำได้ยาก เพราะกรอบตัวอย่างที่มีอยู่มักจะไม่ทันสมัย หรือหากจะสร้างกรอบตัวอย่างขึ้นเอง จะต้องทำการนับจดหรือทำสำมะโน ดังเช่นสำนักงานสถิติแห่งชาติจะมีการทำสำมะโนทุก ๆ 10 ปี โดยจัดทำสำมะโนประชากรและเคหะครั้งแรกใน พ.ศ. 2503 และครั้งล่าสุดใน พ.ศ. 2553 (สำนักงานสถิติแห่งชาติ, 2554) การทำสำมะโนครัวเรือนและเคหะแต่ละครั้งใช้ระยะเวลา งบประมาณ และบุคลากรจำนวนมาก ทำให้ในทางปฏิบัติจำเป็นต้องใช้กรอบตัวอย่างที่ไม่สมบูรณ์ ซึ่งผลการวิเคราะห์อาจมีความเอนเอียงจากค่าที่แท้จริงของประชากร ดังนั้นจำเป็นต้องหาเทคนิคที่ช่วยลดความเอนเอียงนี้ให้น้อยลง เพื่อให้ผลที่ได้นั้นใกล้เคียงกับค่าที่แท้จริงของประชากรมากที่สุด นอกจากนี้ยังพบว่าในการสำรวจมักเกิดปัญหาการไม่ได้รับความร่วมมือ ทำให้ขนาดตัวอย่างมีจำนวนลดลง และส่งผลให้เกิดความคลาดเคลื่อนในการประมวลผลและการวิเคราะห์ข้อมูล และหากหน่วยตัวอย่างเหล่านี้ไม่สามารถลงทะเบียนได้ก็จะส่งผลให้เกิดความเอนเอียงในการวิเคราะห์ข้อมูลด้วย ดังนั้นจะเห็นว่าการใช้กรอบตัวอย่างไม่สมบูรณ์ และการไม่ได้รับความร่วมมือต่างก็ทำให้การประมวลผลและการวิเคราะห์ข้อมูลมีความเอนเอียงไปจากค่าที่แท้จริงของประชากร หากเกิดความคลาดเคลื่อนจากกรอบตัวอย่างและการไม่ได้รับความร่วมมือร่วมกัน ยิ่งจะส่งผลต่อความเอนเอียงอย่างรุนแรงในการประมาณค่าที่แท้จริงของประชากร (Särndal and Lundström, 2005, pp. 179-181) ดังนั้นปัญหากรอบตัวอย่างไม่สมบูรณ์และการไม่ได้รับความร่วมมือจึงมักจะได้รับพิจารณาร่วมกัน

กรอบตัวอย่างที่ไม่สมบูรณ์ มีผลกระทบโดยตรงต่อความน่าจะเป็นแบ่งออกเป็น 3 ลักษณะ คือ 1) กรอบตัวอย่างมีการนับซ้ำ (Duplication) เกิดจากการใช้กรอบตัวอย่างมากกว่า 1 กรอบ ทำให้บางหน่วยของประชากรมีการนับซ้ำ ซึ่งหน่วยที่มีการนับซ้ำจะมีโอกาสถูกเลือกเป็นตัวอย่างมากกว่า 1 ครั้ง 2) กรอบตัวอย่างครอบคลุมมากเกินไป (Overcoverage) เกิดจากการใช้กรอบตัวอย่างที่ครอบคลุมหน่วยที่ไม่ใช่ประชากรเป้าหมาย ทำให้มีโอกาสได้ตัวอย่างที่ไม่ใช่ประชากรเป้าหมาย และ 3) กรอบตัวอย่างไม่ครอบคลุม (Undercoverage) ในกรณีนี้จะเกิดขึ้นเมื่อ

ประชากรที่สนใจศึกษามีขนาดใหญ่กว่ากรอบตัวอย่าง ทำให้กรอบตัวอย่างที่ใช้ไม่นับครอบคลุมทุกหน่วยประชากรที่สนใจศึกษา ตัวอย่างเช่น การสำรวจครัวเรือน หากช่วงเวลาที่จะสำรวจนั้นมีครัวเรือนสร้างที่อยู่อาศัยขึ้นใหม่ ทำให้ครัวเรือนที่สร้างขึ้นใหม่นี้ไม่มีโอกาสถูกเลือกเป็นตัวอย่าง ซึ่งหากดูจากกรอบตัวอย่างแล้วจะไม่ทราบว่า มีหน่วยตัวอย่างบางหน่วยไม่ถูกรวมอยู่ (Non-included) ในกรอบตัวอย่าง หรือมีหน่วยตัวอย่างบางหน่วยที่ปรากฏรายชื่ออยู่ในกรอบตัวอย่างแต่ไม่ใช่ประชากรที่สนใจศึกษา ณ เวลาที่จะสำรวจจะทราบก็ต่อเมื่อมีการสัมภาษณ์หรือมีการแจกบัตร เช่น Andy, et al. (2008) ได้ศึกษาความคลาดเคลื่อนจากการสำรวจทางโทรศัพท์ (Telephone Surveys) เนื่องจากปัญหากรอบตัวอย่างไม่สมบูรณ์และการไม่ได้รับความร่วมมือพบวก่อนการสำรวจจะไม่ทราบว่าเกิดปัญหากรอบตัวอย่างไม่สมบูรณ์ หรือปัญหาการไม่ได้รับความร่วมมือ จนกระทั่งภายหลังการสัมภาษณ์ทางโทรศัพท์ จึงจะสามารถจำแนกได้ว่าหน่วยตัวอย่างใดอยู่หรือไม่อยู่ในกรอบตัวอย่าง และหน่วยตัวอย่างใดให้ความร่วมมือหรือไม่ให้ความร่วมมือ โดยใช้เทคนิคการแบ่งชั้นภูมิภายหลังการแจกบัตร (Post-stratified Techniques) ผลการศึกษาพบว่าการสำรวจทางโทรศัพท์มักเจอปัญหาการไม่ได้รับความร่วมมือมากกว่าปัญหากรอบตัวอย่างไม่สมบูรณ์

การสุ่มตัวอย่างจากกรอบตัวอย่างไม่ครอบคลุมจะพิจารณาให้ U เป็นประชากรเป้าหมาย มีขนาด N_T หน่วย และพบว่าประชากรเป้าหมาย U จะแบ่งออกเป็น 2 ชั้นภูมิ คือ ประชากรที่อยู่ในกรอบตัวอย่าง U_1 (ขนาด N_1 หน่วย) และประชากรที่ไม่อยู่ในกรอบตัวอย่าง U_2 (ขนาด N_2 หน่วย) ดังนั้นจะได้ว่า $U = U_1 \cup U_2$ และ $N_T = N_1 + N_2$ เมื่อเกิดปัญหากรอบตัวอย่างไม่ครอบคลุม พบว่าไม่มีกรอบตัวอย่างสำหรับประชากร U_2 และเกิดปัญหาในการประมาณค่าเฉลี่ยประชากรของตัวแปรที่สนใจศึกษา 3 ประเด็น คือ 1) ไม่มีตัวอย่างสำหรับประชากรที่ไม่อยู่ในกรอบตัวอย่าง 2) ไม่ทราบค่า N_2 และ 3) ไม่ทราบตัวแปรช่วย μ_{x_2} ของประชากร U_2 โดยที่ตัวแปรช่วยจะมีความสัมพันธ์เชิงเส้นตรงกับตัวแปรที่ศึกษา ซึ่งการประมาณค่าที่ใช้ตัวแปรช่วยจะให้การประมาณค่าที่แม่นยำ เมื่อตัวแปรที่ศึกษาและตัวแปรช่วยมีความสัมพันธ์กันสูง

ดังนั้นจากปัญหา 1) ไม่มีตัวอย่างสำหรับประชากรที่ไม่อยู่ในกรอบตัวอย่าง จะแก้ไขปัญหาโดยการสุ่มตัวอย่างแบบง่ายไม่แทนที่ด้วยวิธี Half-open Interval ที่เสนอโดย Hensen, Hurwitz and Madow ปี ค.ศ. 1953 2) ไม่ทราบค่า N_2 จึงประมาณด้วย \hat{N}_2 ที่เสนอโดย Agarwal and Gupta (2008) สำหรับการสุ่มตัวอย่างแบบง่ายไม่แทนที่ เมื่อเกิดปัญหากรอบตัวอย่างไม่ครอบคลุม และ 3) ไม่ทราบสาระค่าเฉลี่ยประชากรของตัวแปรช่วย μ_{x_2} เนื่องจากไม่ทราบ μ_{x_2}

จึงประมาณค่า μ_{x_2} ด้วยค่าเฉลี่ยตัวแปรช่วย μ_{x_1} ของประชากรที่อยู่ในกรอบตัวอย่าง เมื่อ μ_{x_1} และ μ_{x_2} สอดคล้องกัน ดังงานวิจัยของ Ängsved (2004) ที่ศึกษาการประมาณค่าผลรวมประชากรด้วยตัวประมาณการถดถอยภายใต้กรอบตัวอย่างไม่สมบูรณ์

ความคลาดเคลื่อนที่ไม่ได้เกิดจากการสุ่มอีกประเภทหนึ่งที่ควรระมัดระวัง คือ ความคลาดเคลื่อนจากการไม่ได้รับความร่วมมือ ซึ่งเกิดขึ้นจากการไม่สามารถรวบรวมค่าของตัวแปรจากบางหน่วยตัวอย่างด้วยเหตุผลต่าง ๆ เช่น การปฏิเสธที่จะให้ความร่วมมือ การไม่สามารถหาที่อยู่ผู้ให้ข้อมูล ผู้ให้ข้อมูลไม่อยู่บ้านหรือผู้ให้ข้อมูลไม่สามารถตอบคำถามบางข้อได้ แบบสอบถามสูญหายภายหลังการสัมภาษณ์ การจูงใจของพนักงานที่สอบถามไม่ครบทุกหน่วยย่อย เป็นต้น ปัญหาการไม่ได้รับความร่วมมือนี้ทำให้การประมาณค่ามีความเอนเอียงเช่นเดียวกับความคลาดเคลื่อนจากกรอบตัวอย่าง ซึ่งปัญหาการไม่ได้รับความร่วมมือมักหลีกเลี่ยงไม่ได้ ดังนั้นผู้วิจัยจำเป็นต้องอาศัยเทคนิคอื่น ๆ ที่ใช้ในการแก้ไข หรือลดปัญหาการไม่ได้รับความร่วมมือของหน่วยตัวอย่าง (ทวิศักดิ์ ศิริพรไพบูลย์, 2549, หน้า 202-204) โดยอาจใช้วิธีการใดวิธีการหนึ่ง ตัวอย่างเช่น วิธีการสุ่มตัวอย่างย่อยในกลุ่มตัวอย่างที่ไม่ให้ความร่วมมือ (Nonrespondent Subsampling) วิธีการปรับค่าถ่วงน้ำหนัก (Weighting Adjustment Procedures) และการแทนที่ข้อมูลสูญหาย (Imputation Methods) ซึ่งค่าประมาณภายใต้เฉพาะหน่วยที่ให้ความร่วมมือที่จะเป็นค่าประมาณที่ไม่เหมาะสมของค่าประชากร ในทำนองเดียวกันการประมาณค่าที่ใช้ตัวแปรช่วยร่วมในการประมาณค่าก็ไม่เหมาะสมที่จะใช้สำหรับกรณีที่หน่วยที่ไม่ให้ความร่วมมือที่ละทิ้งไม่ได้

วิธีการปรับค่าถ่วงน้ำหนักและการแทนค่าข้อมูลสูญหาย ทั้งสองวิธีมีเป้าหมายในการลดความเอนเอียงจากการไม่ได้รับความร่วมมือ แต่อย่างไรก็ตามวิธีการปรับค่าถ่วงน้ำหนัก และการแทนที่ข้อมูลสูญหาย จะขึ้นอยู่กับข้อตกลงเกี่ยวกับกลไกการให้ความร่วมมือ ซึ่งไม่สามารถควบคุมได้ เมื่อกลไกการให้ความร่วมมือที่สมมตินั้นไม่ถูกต้อง แล้วผลลัพธ์ที่ได้จากการประมาณค่าจะเอนเอียงอย่างรุนแรง ดังนั้นการประมาณค่าที่เหมาะสมเมื่อเกิดการสูญหายแบบไม่สามารถละทิ้งได้ อาจต้องใช้สาระโดยตรงของกลุ่มที่ไม่ให้ความร่วมมือ หรือสาระโดยอ้อมที่เหมาะสมของกลุ่มที่ให้ความร่วมมือในการประมาณค่าสูญหาย ซึ่งแหล่งข้อมูลโดยส่วนใหญ่อันเป็นสาระโดยตรงจากกลุ่มที่ไม่ให้ความร่วมมือก็คือข้อมูลที่ได้จากตัวอย่างของกลุ่มที่ไม่ให้ความร่วมมือนั่นเอง และวิธีการที่จะได้มาของข้อมูลเหล่านั้นก็จะเป็นวิธีการที่ใช้ความพยายามในการเก็บรวบรวมข้อมูลของกลุ่มบุคคลดังกล่าวที่ไม่ให้ความร่วมมือให้กลายเป็นกลุ่มที่ให้ความร่วมมือ เช่น เก็บรวบรวมข้อมูลทางไปรษณีย์ การสัมภาษณ์ทางโทรศัพท์ การสัมภาษณ์ด้วยตนเองในสถานที่ที่ผู้ให้ข้อมูลสะดวกในการให้ความร่วมมือ ซึ่งวิธีการดังกล่าวเสนอโดย Hansen and Hurwitz (1946) ถึงแม้ว่าการสุ่ม

ตัวอย่างย่อยในกลุ่มที่ไม่ให้ความร่วมมือ อาจต้องใช้ค่าใช้จ่ายและระยะเวลาในการดำเนินงานเพิ่มขึ้น กระนั้นถ้าปัญหาความเอนเอียงที่เกิดจากการไม่ได้รับความร่วมมือนี้รุนแรง วิธีการสุ่มตัวอย่างย่อยในกลุ่มที่ไม่ให้ความร่วมมือก็เป็นทางเลือกหนึ่งที่ดำเนินการแก้ไขปัญหาโดยไม่ต้องทำการสำรวจใหม่ให้ได้รับความร่วมมือทั้งหมด อันจะทำให้ต้องเสียเวลาและค่าใช้จ่ายมากกว่า และผลที่ได้จากการสุ่มตัวอย่างย่อยในกลุ่มที่ไม่ให้ความร่วมมือนี้ก็นำมาซึ่งตัวแทนของกลุ่มที่ไม่ให้ความร่วมมือ เพื่อนำไปประมาณค่าตัวแปรที่ศึกษาในกลุ่มที่ไม่ให้ความร่วมมือ เช่น Cochran (1977) ได้เสนอการประมาณค่าเฉลี่ยประชากรด้วยตัวประมาณอย่างง่าย สำหรับการสุ่มตัวอย่างย่อยกลุ่มที่ไม่ให้ความร่วมมือ Okafor and Lee (2000) เสนอการประมาณค่าเฉลี่ยประชากรด้วยตัวประมาณอัตราส่วนและตัวประมาณการถดถอยในแผนการสุ่มตัวอย่างแบบสองขั้นตอน (Double Sampling) เมื่อมีการสุ่มตัวอย่างย่อยในกลุ่มที่ไม่ให้ความร่วมมือ โดยมีเงื่อนไขว่าไม่ทราบตัวแปรช่วยที่คาดว่าจะมีความสัมพันธ์กับตัวแปรที่ศึกษา จึงสุ่มตัวอย่าง 2 ครั้ง โดยครั้งแรกทำการสุ่มตัวอย่างเพื่อศึกษาตัวแปรช่วย จากนั้นจึงสุ่มตัวอย่างครั้งที่สอง ซึ่งสุ่มซ้ำเพียงบางส่วนของตัวอย่างในครั้งแรกเพื่อศึกษาตัวแปรที่ศึกษา และในการสุ่มตัวอย่างครั้งที่สองนี้เกิดปัญหาการไม่ได้รับความร่วมมือจากบางหน่วยของตัวอย่าง จึงทำการสุ่มตัวอย่างย่อยกลุ่มตัวอย่างที่ไม่ให้ความร่วมมืออีกครั้ง และ Siripompibul (2001) ได้ศึกษาตัวประมาณอย่างง่าย เมื่อมีการสุ่มตัวอย่างย่อยในกลุ่มที่ไม่ให้ความร่วมมือ 2 ครั้ง (Nonrespondent Subsampling Two Scheme)

ดังนั้นจากปัญหากรอบตัวอย่างไม่ครอบคลุมและการไม่ได้รับความร่วมมือ และจากการศึกษางานวิจัยข้างต้น ผู้วิจัยจึงสนใจศึกษา 1) การประมาณค่าเฉลี่ยประชากร เมื่อเกิดปัญหากรอบตัวอย่างไม่ครอบคลุม 2) การประมาณค่าเฉลี่ยประชากร เมื่อเกิดปัญหาการไม่ได้รับความร่วมมือ และ 3) การประมาณค่าเฉลี่ยประชากร เมื่อเกิดปัญหากรอบตัวอย่างไม่ครอบคลุมและการไม่ได้รับความร่วมมือร่วมกัน ซึ่งศึกษากรณีทีภายหลังการแจกแจงนับไม่มีทะเบียนใหม่ล่าสุด และไม่มีกรอบตัวอย่างสำหรับประชากรที่ไม่อยู่ในกรอบตัวอย่าง จึงทำให้ไม่ทราบค่า N_2 และ μ_{x_2} ผู้วิจัยจึงประมาณค่า N_2 ดังงานวิจัยของ Agarwal and Gupta (2008) และเมื่อ μ_{x_1} และ μ_{x_2} สอดคล้องกัน จะประมาณค่าเฉลี่ยประชากรตัวแปรช่วย μ_{x_2} ด้วยค่าเฉลี่ยตัวแปรช่วย μ_{x_1} ของประชากรที่อยู่ในกรอบตัวอย่าง แต่เพื่อตรวจสอบว่ากรณี μ_{x_1} และ μ_{x_2} ต่างกัน จะมีอิทธิพลต่อการประมาณค่าหรือไม่ จึงศึกษาสถานการณ์ที่ μ_{x_1} และ μ_{x_2} สอดคล้องกัน และ μ_{x_1} และ μ_{x_2} ต่างกัน และในการแก้ไขปัญหาที่ไม่ได้ตัวอย่างสำหรับประชากรที่ไม่อยู่ในกรอบตัวอย่าง จึงสุ่มตัวอย่างแบบง่ายไม่แทนที่ เมื่อเกิดปัญหากรอบตัวอย่างไม่ครอบคลุมด้วยวิธี Half-open Interval ที่เสนอโดย Hensen, Hurwitz and Madow และดำเนินเก็บข้อมูลจากหน่วยตัวอย่างที่ไม่อยู่ใน

กรอบตัวอย่าง คือ ค่าของตัวที่ศึกษา ค่าของตัวแปรช่วย และจำนวนหน่วยตัวอย่างที่ไม่อยู่ในกรอบตัวอย่าง และผู้วิจัยได้เสนอการประมาณค่าเฉลี่ยประชากรโดยใช้วิธีการแบ่งชั้นภูมิภายหลังการแจกแจง ซึ่งแยกพิจารณา 2 กรณี คือ 1) กรณีประชากรที่อยู่ในกรอบตัวอย่างและประชากรที่ไม่อยู่ในกรอบตัวอย่างมีคุณลักษณะของค่าเฉลี่ยสอดคล้องกัน และ 2) กรณีประชากรที่อยู่ในกรอบตัวอย่าง และประชากรที่ไม่อยู่ในกรอบตัวอย่างมีคุณลักษณะของค่าเฉลี่ยต่างกัน ด้วยค่าเฉลี่ยตัวอย่างที่ได้จากกรอบตัวอย่าง และตัวอย่างที่ไม่อยู่ในกรอบตัวอย่างร่วมกับการใช้ค่าประมาณจำนวนประชากรที่ไม่อยู่ในกรอบตัวอย่าง และโดยทั่วไปมักไม่ทราบกลไกการไม่ให้ความร่วมมือ จึงต้องใช้สาระโดยตรงจากกลุ่มที่ไม่ให้ความร่วมมือ ด้วยการสุ่มย่อยในกลุ่มที่ไม่ให้ความร่วมมือ เพื่อเปลี่ยนผู้ที่ไม่ให้ความร่วมมือให้กลายเป็นผู้ให้ความร่วมมือ และประมาณค่าเฉลี่ยด้วยตัวประมาณอย่างง่าย ตัวประมาณอัตราส่วน และตัวประมาณการถดถอย และเปรียบเทียบประสิทธิภาพของตัวประมาณในแต่ละสถานการณ์ เพื่อใช้ในการพิจารณาเลือกใช้ตัวประมาณให้เหมาะสมกับสถานการณ์ต่าง ๆ เมื่อเกิดปัญหากรอบตัวอย่างไม่ครอบคลุมและการไม่ได้รับความร่วมมือ

จุดมุ่งหมายของการวิจัย

1. เพื่อเปรียบเทียบประสิทธิภาพตัวประมาณอย่างง่าย (Simple Estimator) ตัวประมาณอัตราส่วน (Ratio Estimator) และตัวประมาณการถดถอย (Regression Estimator) ในแผนการสุ่มตัวอย่างแบบง่าย เมื่อเกิดปัญหากรอบตัวอย่างไม่ครอบคลุม
2. เพื่อเปรียบเทียบประสิทธิภาพตัวประมาณอย่างง่าย ตัวประมาณอัตราส่วน และตัวประมาณการถดถอย ในแผนการสุ่มตัวอย่างแบบง่าย เมื่อเกิดปัญหาการไม่ได้รับความร่วมมือ
3. เพื่อเปรียบเทียบประสิทธิภาพตัวประมาณอย่างง่าย ตัวประมาณอัตราส่วน และตัวประมาณการถดถอย เมื่อเกิดปัญหากรอบตัวอย่างไม่ครอบคลุมและการไม่ได้รับความร่วมมือ

ขอบเขตของการวิจัย

ในการวิจัยครั้งนี้ ผู้วิจัยได้กำหนดขอบเขตของการวิจัยดังนี้

1. ประชากรที่ใช้ในการศึกษามี 2 ปัจจัยหลัก ๆ ที่กำหนดสถานการณ์การศึกษา ได้แก่ ปัจจัยกรอบตัวอย่างไม่ครอบคลุม และปัจจัยการไม่ได้รับความร่วมมือ
 - 1.1 กำหนดขนาดประชากรเท่ากับ 1,000
 - 1.2 กำหนดสัดส่วนประชากรที่ไม่อยู่ในกรอบตัวอย่าง 4 ระดับ คือ 5%, 10%, 20% และ 30% ของประชากร

1.3 กำหนดสัดส่วนการไม่ได้รับความร่วมมือ 4 ระดับ คือ 5%, 10%, 20% และ 30% ของประชากรที่อยู่ในกรอบตัวอย่าง

1.4 กำหนดประชากรที่อยู่ในกรอบตัวอย่าง และประชากรที่ไม่อยู่ในกรอบตัวอย่าง มีคุณลักษณะของค่าเฉลี่ยสอดคล้องกัน ($\mu_1 = \mu_2$) และมีคุณลักษณะของค่าเฉลี่ยต่างกัน ($\mu_1 \neq \mu_2$)

1.5 กำหนดประชากรที่อยู่ในกรอบตัวอย่าง และประชากรที่ไม่อยู่ในกรอบตัวอย่าง มีคุณลักษณะของค่าเฉลี่ยตัวแปรช่วยสอดคล้องกัน ($\mu_{x_1} = \mu_{x_2}$) และมีคุณลักษณะของค่าเฉลี่ยตัวแปรช่วยต่างกัน ($\mu_{x_1} \neq \mu_{x_2}$)

2. ตัวแปรที่ใช้ในการศึกษาแต่ละสถานการณ์ประกอบด้วย

2.1 ตัวแปร Y คือ ตัวแปรที่ศึกษาเป็นข้อมูลเชิงปริมาณ

2.2 ตัวแปร X คือ ตัวแปรช่วยเป็นข้อมูลเชิงปริมาณที่มีความสัมพันธ์เชิงเส้นตรงกับตัวแปร Y โดยที่ความสัมพันธ์แบ่งออกเป็น 3 ระดับ และแบ่งช่วงของค่าสัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน (Pearson's Correlation Coefficient) ดังนี้

2.2.1 ความสัมพันธ์ระดับต่ำ อยู่ในช่วง (0, 0.4)

2.2.2 ความสัมพันธ์ระดับปานกลาง อยู่ในช่วง [0.4, 0.7)

2.2.3 ความสัมพันธ์ระดับสูง อยู่ในช่วง [0.7, 1)

3. ใช้แผนการสุ่มตัวอย่างแบบง่ายไม่แทนที่ในการสุ่มตัวอย่าง และกำหนดขนาดตัวอย่างที่ต้องการศึกษา 10%, 20% และ 30% ของประชากรในกรอบตัวอย่าง

4. ประมาณค่าเฉลี่ยประชากรด้วยตัวประมาณ 3 วิธี คือ ตัวประมาณอย่างง่าย ตัวประมาณอัตราส่วน และตัวประมาณการถดถอย

5. ในแต่ละสถานการณ์จะทำซ้ำ 10,000 รอบ และตัดสินใจโดยใช้เกณฑ์ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุด

ข้อตกลงเบื้องต้น

1. การสุ่มตัวอย่างแบบง่ายไม่แทนที่ เมื่อเกิดปัญหากรอบตัวอย่างไม่ครอบคลุม จะกระทำได้อีกต่อเมื่อกรอบตัวอย่างที่ใช้ในการสุ่มตัวอย่างนั้น เป็นกรอบตัวอย่างที่ระบุตำแหน่งที่ตั้งหรือที่อยู่ของหน่วยตัวอย่างที่ชัดเจนและแน่นอน เช่น คริวเรือ นสถานประกอบการ เป็นต้น และในบางครั้งกรอบตัวอย่างที่ใช้นั้นจำเป็นต้องมีแผนที่สำหรับใช้ในการระบุตำแหน่งที่ตั้งของหน่วยตัวอย่างด้วย และหน่วยประชากรที่ไม่อยู่ในกรอบตัวอย่างกระจายทั่วไปในประชากร ดังนั้นจึง

สามารถหาที่อยู่ หรือหาตำแหน่งของหน่วยตัวอย่างที่อยู่ในกรอบตัวอย่างและหน่วยตัวอย่างที่ไม่อยู่ในกรอบตัวอย่างได้ด้วยวิธี Half-open Interval

2. การวิจัยครั้งนี้ศึกษาปัญหาการไม่ได้รับความร่วมมือบางหน่วยในตัวอย่างที่อยู่ในกรอบตัวอย่าง และศึกษากรณีที่กรอบตัวอย่างที่ใช้ในการสำรวจนั้นมีการระบุตำแหน่งที่ตั้งของหน่วยตัวอย่างที่ชัดเจน จึงทำให้สามารถหาที่อยู่ หรือหน่วยตัวอย่างที่อยู่ในกรอบตัวอย่าง และหาหน่วยตัวอย่างที่ไม่อยู่ในกรอบตัวอย่างได้ ดังนั้นการไม่ได้รับความร่วมมือบางหน่วย อาจเกิดขึ้นจากการไม่สามารถรวบรวมค่าของตัวแปรจากบางหน่วยตัวอย่างด้วยเหตุผลต่าง ๆ เช่น การปฏิเสธที่จะให้ความร่วมมือ ผู้ให้ข้อมูลไม่อยู่บ้าน แบบสอบถามสูญหายภายหลังการสัมภาษณ์ เป็นต้น

ประโยชน์ที่คาดว่าจะได้รับ

1. ทราบประสิทธิภาพของตัวประมาณค่าเฉลี่ย ภายใต้สถานการณ์ที่ศึกษา และเป็นแนวทางเลือกตัวประมาณค่าเฉลี่ย เมื่อเกิดปัญหากรอบตัวอย่างไม่ครอบคลุมหรือปัญหาการไม่ได้รับความร่วมมือ

2. เป็นแนวทางในการนำไปประยุกต์ใช้เมื่อเกิดปัญหากรอบตัวอย่างไม่ครอบคลุม หรือการไม่ได้รับความร่วมมือ

3. เป็นแนวทางในการประมาณค่าเฉลี่ย เมื่อมีตัวแปรช่วยหรือไม่มีตัวแปรช่วยในกรอบตัวอย่างสำหรับสถานการณ์ต่าง ๆ เพื่อลดค่าความเอนเอียงเมื่อเกิดปัญหากรอบตัวอย่างไม่ครอบคลุมหรือปัญหาการไม่ได้รับความร่วมมือ

4. เป็นแนวทางในการศึกษาวิธีประมาณค่าเมื่อเกิดปัญหากรอบตัวอย่างไม่ครอบคลุมในสถานการณ์อื่น ๆ รวมทั้งแนวทางการแก้ปัญหาการไม่ได้รับความร่วมมือ