บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

การวิจัยนี้มีทฤษฎี เอกสาร และงานวิจัยที่เกี่ยวข้องกับการเปรียบเทียบประสิทธิภาพตัวประมาณค่าเฉลี่ย คือ ตัวประมาณอย่างง่าย ตัวประมาณอัตราส่วน และตัวประมาณการถดถอยในแผนการสุ่มตัวอย่างแบบง่ายไม่แทนที่ เมื่อเกิดปัญหากรอบตัวอย่างไม่ครอบคลุม และการไม่ได้รับความร่วมมือ ดังนี้

1. การสำรวจด้วยตัวอย่าง

    1.1 วัตถุประสงค์ของการสำรวจด้วยตัวอย่าง

    1.2 การสุ่มตัวอย่างโดยใช้ความน่าจะเป็น

    1.3 ความคลาดเคลื่อนที่ไม่ได้เกิดจากการสุ่มตัวอย่าง

2. กรอบตัวอย่าง

    2.1 กรอบตัวอย่างที่สมบูรณ์

    2.2 กรอบตัวอย่างไม่สมบูรณ์

    2.3 กรอบตัวอย่างไม่ครอบคลุม

    2.4 การสุ่มตัวอย่างแบบง่ายไม่แทนที่ และการประมาณค่าเฉลี่ยประชากร เมื่อเกิดปัญหากรอบตัวอย่างไม่ครอบคลุม

3. การไม่ได้รับความร่วมมือ

    3.1 ความคลาดเคลื่อนจากการไม่ได้รับความร่วมมือ

    3.2 แนวทางแก้ไขปัญหาการไม่ได้รับความร่วมมือ

    3.3 การสุ่มตัวอย่างย่อยในกลุ่มที่ไม่ให้ความร่วมมือ

    3.4 การประมาณค่าเฉลี่ยประชากร เมื่อเกิดปัญหาการไม่ได้รับความร่วมมือด้วยการสุ่มตัวอย่างย่อยในกลุ่มที่ไม่ให้ความร่วมมือ

4. การประมาณค่าเฉลี่ยประชากร เมื่อเกิดปัญหากรอบตัวอย่างไม่ครอบคลุม และการไม่ได้รับความร่วมมือ

5. การเปรียบเทียบประสิทธิภาพของตัวประมาณ

6. งานวิจัยที่เกี่ยวข้อง

การสำรวจด้วยตัวอย่าง

การสำรวจด้วยตัวอย่าง เป็นการเก็บรวบรวมข้อมูลจากตัวอย่าง เพื่อนำมาวิเคราะห์ และสรุปผลเกี่ยวกับคุณลักษณะของประชากร ตัวอย่างที่ใช้ควรเป็นตัวอย่างสุ่ม และมีขนาดเหมาะสม เลือกจากกรอบตัวอย่างที่สมบูรณ์ อาจเป็นกรอบรายชื่อหรือกรอบพื้นที่ ซึ่งการสำรวจด้วยตัวอย่างจะให้ประสิทธิภาพสูงต้องอาศัยการออกแบบแผนเลือกตัวอย่างที่ดี ใช้วิธีประมวลผลและการวิเคราะห์ข้อมูลที่เหมาะสม โดยลักษณะการสำรวจด้วยตัวอย่างที่ดีต้องมีการระบุประชากรที่ชัดเจนก่อนการสำรวจ จากนั้นต้องหาหรือสร้างกรอบตัวอย่างที่ดีและทันสมัย และเลือกใช้แผนการสุ่มให้เหมาะสมกับคุณลักษณะของประชากรในกรอบตัวอย่าง นอกจากนี้ยังต้องคำนึงถึงข้อจำกัดต่าง ๆ สำหรับการวางแผนการสำรวจ เช่น เวลา ค่าใช้จ่าย บุคลากร การประมวลผล และการวิเคราะห์ข้อมูล เพื่อให้แผนการสำรวจอยู่ในวิสัยที่สามารถปฏิบัติได้จริง โดยมีรายละเอียดเกี่ยวกับการสำรวจด้วยตัวอย่าง ดังนี้

1. **วัตถุประสงค์ของการสำรวจด้วยตัวอย่าง** (Objective of Sampling Survey)

การสำรวจด้วยตัวอย่างมีวัตถุประสงค์เพื่อเก็บรวบรวมข้อมูล (ค่าของตัวแปร) จากข้อมูลที่มีอยู่แล้วหรือเกิดขึ้นตามธรรมชาติ โดยการเก็บรวบรวมข้อมูลหรือการแจงนับบางหน่วยของประชากร ที่เรียกว่า "ตัวอย่าง" อาจทำด้วยการสอบถาม การสัมภาษณ์ การสังเกตการณ์ การวัดโดยตรง หรือการทอดแบบ รวมทั้งการนำข้อมูลทุติยภูมิที่มีผู้เก็บรวบรวมข้อมูลแล้วในรูปของรายงาน แฟ้มข้อมูล ระเบียน ฯลฯ เพื่อนำมาประมวลผลหรือวิเคราะห์ต่อไป

ในการสำรวจจะเรียกหน่วยที่ให้ข้อมูลว่า "หน่วยตัวอย่าง" (Sampling Unit) ซึ่งอาจจะเป็นบุคคล หรือกลุ่มบุคคล เช่น ครัวเรือน หมู่บ้าน ตำบล ฯลฯ นอกจากนั้นอาจเป็นหน่วยงาน เช่น บริษัท โรงงาน สถานบริการ หรือเป็นพื้นที่ เช่น พื้นที่ป่า พื้นที่เพาะปลูก เป็นต้น โดยที่หน่วยตัวอย่างจะเป็นอะไรขึ้นอยู่กับแผนการสุ่มตัวอย่าง (Sampling Design) ที่ใช้ และตัวแปรที่ศึกษา

หน่วยตัวอย่างทั้งหมดที่ครอบคลุมในการศึกษาเรียกว่าประชากร (Population) หรือประชากรเป้าหมาย และจำนวนหน่วยของตัวอย่างทั้งหมดในประชากรเรียกว่า ขนาดประชากร (Population Size) มักแทนด้วย N และจะเรียกส่วนหนึ่งของประชากรว่า "ตัวอย่าง" และจำนวนหน่วยในตัวอย่างจะเรียกว่า "ขนาดตัวอย่าง (Sample Size)" และมักแทนด้วย n ตัวอย่างสุ่ม (Random Sample) ได้แก่ ตัวอย่างที่ได้จากประชากรหนึ่งด้วยวิธีการที่สามารถบอกได้ว่าตัวอย่างนั้นจะมีโอกาสถูกเลือกให้มาอยู่ในตัวอย่างมากน้อยเพียงไร ซึ่งในทางปฏิบัติจะมีวิธีการเลือกมาโดยการสุ่ม (Randomization) หรือ**การสุ่มตัวอย่างโดยใช้ความน่าจะเป็น (Probability Sampling)** เช่น การจับสลาก หรือการใช้ตารางเลขสุ่ม เป็นต้น ซึ่งตัวอย่างที่ได้โดยการสุ่มนี้ถือได้

ว่าเป็นตัวแทน (Representative) ของประชากร โดยที่ตัวอย่างสุ่มนี้สามารถทำการอนุมานไปยังประชากรได้ ส่วนตัวอย่างที่ไม่ใช่ตัวอย่างสุ่ม (Nonrandom Sample) หรือ**การสุ่มตัวอย่างโดยไม่ใช้ความน่าจะเป็น (Non-probability Sampling)** ซึ่งเป็นการสุ่มตัวอย่างที่ไม่สามารถบอกได้ว่าหน่วยตัวอย่างของประชากรจะถูกเลือกให้อยู่ในตัวอย่างด้วยความน่าจะเป็นเท่าใด จึงไม่อาจขยายความหรืออนุมานไปยังประชากรได้ ดังนั้นหากต้องการศึกษาคุณลักษณะของประชากร จึงจำเป็นที่ต้องสุ่มตัวอย่างโดยใช้ความน่าจะเป็น (ประชุม สุวัตถี, 2552, หน้า 4-5)

2. **การสุ่มตัวอย่างโดยใช้ความน่าจะเป็น**

การสุ่มตัวอย่างโดยใช้ความน่าจะเป็นคือการเลือกตัวอย่างจากประชากรที่แต่ละหน่วยตัวอย่างมีความน่าจะเป็นที่ถูกเลือกตามที่กำหนดขึ้น โดยที่ความน่าจะเป็นของการสุ่มจะขึ้นอยู่กับกระบวนการสุ่มหรือแผนการสุ่มตัวอย่าง การสุ่มตัวอย่างโดยใช้ความน่าจะเป็นแบ่งออกเป็น 2 ประเภท ดังนี้

2.1 การสุ่มตัวอย่างโดยใช้ความน่าจะเป็นไม่เท่ากัน (Unequal Probability Sampling) การสุ่มตัวอย่างโดยใช้ความน่าจะเป็นไม่เท่ากันเป็นขบวนการเลือกตัวอย่าง ซึ่งหน่วยตัวอย่างในแต่ละหน่วยประชากรมีความน่าจะเป็นที่จะอยู่ในตัวอย่างที่ต้องการเลือกมาศึกษาด้วยค่าที่ไม่เท่ากัน ซึ่งความน่าจะเป็นดังกล่าวแฝงอยู่ในขั้นตอนการเลือกตัวอย่าง ดังนั้นโอกาสที่หน่วยตัวอย่างถูกเลือกยิ่งสูงมากขึ้นเท่าใด ก็ย่อมแสดงว่าหน่วยตัวอย่างนั้นค่อนข้างสำคัญ และมีอิทธิพลต่อการประมาณค่า เช่น ในการศึกษาเกี่ยวกับป่าไม้ในพื้นที่แห่งหนึ่ง ถ้าต้นไม้เขตไหนมีจำนวนมากโอกาสที่จะได้รับการคัดเลือกในพื้นที่นั้นก็จะมากไปด้วย หรือการศึกษาผลผลิตของการปลูกข้าวถ้าครัวเรือนไหนมีพื้นที่ทำการเกษตรมาก ครัวเรือนนั้นก็จะมีโอกาสถูกคัดเลือกเป็นตัวอย่างมากกว่าครัวเรือนที่มีพื้นที่ทำการเกษตรน้อย

2.2 การสุ่มตัวอย่างโดยใช้ความน่าจะเป็นเท่ากัน (Equal Probability Sampling) การสุ่มตัวอย่างโดยใช้ความน่าจะเป็นเท่ากันเป็นขบวนการเลือกตัวอย่างที่หน่วยตัวอย่างในแต่ละหน่วยประชากรจะมีความน่าจะเป็นที่จะอยู่ในตัวอย่างที่ต้องการเลือกมาศึกษาด้วยค่าที่เท่ากันซึ่งความน่าจะเป็นดังกล่าวแฝงอยู่ในขั้นตอนการเลือกตัวอย่างหรือ**แผนการสุ่มตัวอย่าง** ดังนั้นโอกาสที่หน่วยตัวอย่างถูกเลือกนั้นจะเท่ากัน นั่นแสดงว่าหน่วยตัวอย่างแต่ละหน่วยนั้นมีความสำคัญเท่ากัน โดยกระบวนการในการเลือกตัวอย่างโดยใช้ความน่าจะเป็น แบ่งออกเป็น 4 แผนการสุ่มตัวอย่าง ดังนี้

2.2.1 การสุ่มตัวอย่างแบบง่าย (Simple Sampling) เป็นการเลือกตัวอย่างทีละหน่วยด้วยความน่าจะเป็นเท่ากัน ๆ ซึ่งมักจะใช้เลขสุ่มหรือการจับสลาก อาจเป็นการสุ่มแบบแทนที่ (Sampling With Replacement) หรือแบบไม่แทนที่ (Sampling Without Replacement)

โดยทั่วไปมักใช้การสุ่มแบบไม่แทนที่ ซึ่งแผนการสุ่มตัวอย่างแบบง่ายเหมาะสมกับประชากรที่มีคุณลักษณะคล้ายคลึงกัน

2.2.2 การสุ่มตัวอย่างแบบมีระบบ (Systematic Sampling) เป็นการสุ่มตัวอย่างหนึ่งหน่วยจากทุก ๆ a หน่วย โดยมีการเรียงลำดับหน่วยตัวอย่างในกรอบตัวอย่างหน่วยที่ 1, 2,..., N ถ้าต้องการสุ่ม n หน่วย จะกำหนดให้ a เป็นจำนวนเต็มที่ใกล้ N/n ที่สุด ซึ่งจะทำให้ตัวอย่างแต่ละตัวถูกเลือกด้วยความน่าจะเป็นเท่า ๆ กัน

2.2.3 การสุ่มตัวอย่างแบบแบ่งชั้นภูมิ (Stratified Sampling) เป็นการเลือกตัวอย่างที่มีการแบ่งประชากรออกเป็นเซตย่อยที่ไม่มีส่วนร่วมกันเรียกว่า "ชั้นภูมิ" แล้วสุ่มตัวอย่างในแต่ละชั้นภูมิด้วยวิธีใดวิธีหนึ่ง เช่น ถ้าสุ่มตัวอย่างแบบง่ายก็จะเรียกว่า การสุ่มตัวอย่างแบบแบ่งชั้นภูมิอย่างง่าย (Stratified Random Sampling) และถ้าในชั้นภูมินั้นทำการสุ่มแบบมีระบบจะเรียกว่าการสุ่มตัวอย่างแบบแบ่งชั้นภูมิอย่างมีระบบ (Stratified Systematic Sampling) การสุ่มตัวอย่างแบบแบ่งชั้นภูมินี้จะให้การประมาณพารามิเตอร์แม่นยำขึ้น หรือต้องการข้อเท็จจริงในส่วนต่าง ๆ ของประชากร แต่การจะได้ผลตามที่ต้องการหรือไม่ต้องขึ้นอยู่กับความเหมาะสมในการแบ่งชั้นภูมิ การแบ่งชั้นภูมิที่ดีคือประชากรที่อยู่ในชั้นภูมิเดียวกันจะมีคุณลักษณะคล้ายคลึงกัน แต่ต่างชั้นภูมิจะมีคุณลักษณะที่แตกต่างกัน

2.2.4 การสุ่มตัวอย่างแบบกลุ่ม (Cluster Sampling) เป็นวิธีการสุ่มตัวอย่างที่จะสุ่มมารวมอยู่ในตัวอย่างเป็นกลุ่มของหน่วยเล็กหลายหน่วยของประชากร ซึ่งแต่ละหน่วยเล็กของประชากรมีคุณลักษณะคล้ายคลึงกัน เราจะเลือกบางหน่วยของประชากรเพื่อเป็นกลุ่มประชากรตัวอย่าง จากนั้นก็ทำการเลือกตัวอย่างในกลุ่มประชากรตัวอย่างนั้นด้วยวิธีการสุ่มอย่างแบบง่ายหรือการสุ่มตัวอย่างแบบมีระบบ

การเลือกแผนการสุ่มตัวอย่างขึ้นอยู่กับคุณลักษณะของประชากร เช่น ถ้าประชากรในกรอบตัวอย่างมีคุณลักษณะสอดคล้องกัน แผนการสุ่มตัวอย่างที่เหมาะสมก็คือแผนการสุ่มตัวอย่างแบบง่าย และนอกจากนี้การประมาณค่าประชากรที่สนใจศึกษา หรือที่เรียกว่าค่าพารามิเตอร์ (Parameters) ให้มีความแม่นยำ จำเป็นต้องใช้วิธีการประมาณค่าที่เหมาะสมด้วย เช่น เมื่อต้องการประมาณค่าพารามิเตอร์ของตัวแปรที่ศึกษา (Study Variable) อาจประมาณค่าพารามิเตอร์ด้วยวิธีการประมาณค่าอย่างง่าย (Simple Estimator Method) หรือวิธีการประมาณค่าที่ใช้ตัวแปรที่เกี่ยวข้องหรือตัวแปรที่มีความสัมพันธ์กับตัวแปรที่ศึกษา (Composite Estimator Method) ซึ่งตัวแปรดังกล่าวอาจเรียกว่า ตัวแปรช่วย (Auxiliary Variables) โดยวิธี

ประมาณค่าที่นิยมใช้ คือ วิธีประมาณค่าอัตราส่วน (Ratio Estimator Method) และวิธีประมาณค่าการถดถอย (Regression Estimator Method)

3. **ความคลาดเคลื่อนที่ไม่ได้เกิดจากการสุ่มตัวอย่าง**

ความคลาดเคลื่อนที่ไม่ได้เกิดจากการสุ่มตัวอย่าง มักเกิดจากการใช้นิยามศัพท์ และแนวคิดไม่ถูกต้องหรือไม่เหมาะสม การใช้กรอบตัวอย่างที่ไม่สมบูรณ์ (Imperfect Frame) การใช้แบบสอบถามบกพร่อง การใช้วิธีการเก็บรวบรวมข้อมูลผิดพลาด การลงรหัส หรือใช้สถิติที่ผิด ซึ่งความคลาดเคลื่อนที่ไม่ได้เกิดจากการสุ่มตัวอย่างมักมีรูปแบบไม่ชัดเจน และควบคุมได้ยาก มีขนาดใหญ่ขึ้นเมื่อขนาดตัวอย่างเพิ่มขึ้น ดังนั้นหากไม่ควบคุมให้ดีความคลาดเคลื่อนที่ไม่ได้เกิดจากการสุ่มตัวอย่างจะทำความเสียหายต่อการสำรวจมากกว่าความคลาดเคลื่อนที่เกิดจากการสุ่มตัวอย่าง โดยเฉพาะอย่างยิ่งในการสำรวจครัวเรือนที่มีขนาดตัวอย่างใหญ่มาก ๆ

ความคลาดเคลื่อนที่ไม่เกิดจากการสุ่มตัวอย่าง จะทำให้เกิดความเอนเอียง (Biased) ของตัวประมาณ โดยความเอนเอียงเป็นผลต่างระหว่างค่าประมาณกับค่าพารามิเตอร์ ซึ่งเป็นส่วนหนึ่งของความคลาดเคลื่อนที่เกิดจากการสุ่ม แต่ความเอนเอียงที่ไม่เกิดจากการสุ่มตัวอย่าง จะเป็นความเอนเอียงที่มาจากการเลือกตัวอย่างไม่เหมาะสม ทำให้คำตอบไม่ตรงกับความจริงหรือไม่ได้คำตอบเลย หรืออาจเก็บข้อมูลจากหน่วยที่ไม่ใช่ประชากรเป้าหมาย ดังนั้นความคลาดเคลื่อนที่ไม่ได้เกิดจากการสุ่มตัวอย่าง จะส่งผลต่อความเอนเอียงในการประมาณค่ามากกว่าความคลาดเคลื่อนที่เกิดจากการสุ่มตัวอย่าง โดยที่แหล่งของความคลาดเคลื่อนที่ไม่ได้เกิดจากการสุ่มตัวอย่างมีดังนี้

3.1 ความคลาดเคลื่อนคุ้มรวม (Coverage Errors) เกิดขึ้นเมื่อเส้นกั้นอาณาเขตของหน่วยตัวอย่างไม่ชัดเจนพอ หรือเกิดจากการใช้กรอบตัวอย่างไม่สมบูรณ์ ทำให้หน่วยตัวอย่าง เช่น ชุมรุมอาคาร หมู่บ้าน ครัวเรือนที่ซ้ำซ้อนกัน หรือไม่ครอบคลุมครัวเรือนครบถ้วนตามที่ต้องการ หรือครอบคลุมเกินขอบเขตที่ต้องการศึกษา

3.2 ความคลาดเคลื่อนจากการไม่ได้รับความร่วมมือ (Nonresponse Errors) เกิดจากการไม่สามารถสังเกต หรือรวบรวมค่าของตัวแปรจากบางหน่วยตัวอย่างด้วยเหตุผลต่าง ๆ เช่น การปฏิเสธที่จะให้ความร่วมมือ การไม่สามารถหาที่อยู่ของตัวอย่างได้ หรือผู้ตอบไม่อยู่บ้าน หรือผู้ตอบไม่สามารถตอบคำถามบางข้อได้ แบบสอบถามสูญหายภายหลังการสัมภาษณ์ การจงใจของพนักงานที่สอบถามไม่ครบทุกหน่วยย่อย ฯลฯ

3.3 ความคลาดเคลื่อนจากการวัด (Measurement Errors) เกิดขึ้นจากค่าสังเกตหรือค่าที่วัดได้ต่างจากค่าที่แท้จริง ความคลาดเคลื่อนประเภทนี้มักเกิดจากกระบวนการสำรวจ เช่น วัตถุประสงค์ของการสำรวจ การตั้งคำถาม การสัมภาษณ์ การบันทึกคำตอบ การลงรหัส

การประมวลผลข้อมูล และการรายงานผลการสำรวจ ซึ่งเป็นความคลาดเคลื่อนที่เกิดได้จากแต่ละหน่วยสังเกต ความคลาดเคลื่อนนี้อาจเกิดจากตัวผู้ให้คำตอบที่ไม่เข้าใจคำถาม หรือพยายามให้ความร่วมมือทั้งที่ไม่ทราบจะตอบอย่างไร หรือตั้งใจให้คำตอบผิด หรือเกิดจากตัวผู้สอบถามที่ไม่ทราบวัตถุประสงค์ของการสำรวจซึ่งถามด้วยคำถามที่ไม่ชัดเจน หรือการละเลยหน้าที่ที่จะสอบถาม เป็นต้น

3.4 ความคลาดเคลื่อนจากการประมวลผล (Processing Errors) เกิดขึ้นได้ทุกขั้นตอนของการดำเนินการประมวลผลข้อมูล เช่น ความความเคลื่อนในการทำบรรณาธิการ การลงรหัส การนำเข้าข้อมูล การใช้โปรแกรม ฯลฯ

ความคลาดเคลื่อนที่ไม่ได้เกิดจากการสุ่มตัวอย่างที่เกิดจากการประมวลผล หรือการวัดที่ผิดพลาด สามารถลดความคลาดเคลื่อนนี้ด้วยการอบรมพนักงานที่เกี่ยวข้องให้มีความเชี่ยวชาญในการแจงนับ การวัด การลงรหัสข้อมูล ตลอดจนการตรวจสอบความถูกต้องของข้อมูลก่อนการลงรหัสข้อมูล ส่วนความคลาดเคลื่อนที่เกิดจากกรอบตัวอย่างเป็นความคลาดเคลื่อนที่ควบคุมและหลีกเลี่ยงได้ยาก เนื่องมาจากการหาหรือสร้างกรอบตัวอย่างที่สมบูรณ์นั้นทำได้ยาก เพราะกรอบตัวอย่างที่มีอยู่มักจะไม่ทันสมัยหรือหากจะสร้างกรอบตัวอย่างขึ้นเอง จะต้องทำการนับจดหรือทำสำมะโน (Census) ดังเช่นสำนักงานสถิติแห่งชาติจะมีการทำสำมะโนทุก ๆ 10 ปี โดยจัดทำสำมะโนประชากรและเคหะครั้งแรกใน พ.ศ. 2503 และครั้งล่าสุดใน พ.ศ. 2553 (สำนักงานสถิติแห่งชาติ, 2554) การทำสำมะโนครัวเรือน และเคหะแต่ละครั้งใช้ระยะเวลา งบประมาณ และบุคลากรจำนวนมาก ทำให้ในทางปฏิบัติจำเป็นต้องใช้กรอบตัวอย่างที่ไม่สมบูรณ์ ซึ่งผลที่ได้จากการวิเคราะห์จะมีความเอนเอียงจากค่าที่แท้จริงของประชากร ดังนั้นจำเป็นต้องหาเทคนิคที่ช่วยลดความเอนเอียงนี้ให้น้อยลง เพื่อให้ผลที่ได้นั้นใกล้เคียงกับค่าพารามิเตอร์มากที่สุด นอกจากนี้ยังพบว่าในการสำรวจมักเกิดปัญหาการไม่ได้รับความร่วมมือ ทำให้ขนาดตัวอย่างมีจำนวนลดลง และเกิดความคลาดเคลื่อนในการประมวลผล และการวิเคราะห์ข้อมูล และหากหน่วยตัวอย่างเหล่านี้ไม่สามารถละทิ้งได้ก็จะส่งผลให้เกิดความเอนเอียงในการวิเคราะห์ข้อมูลด้วย

กรอบตัวอย่าง

ความคลาดเคลื่อนที่เกิดจากกรอบตัวอย่าง เป็นความคลาดเคลื่อนที่ไม่ได้เกิดจากการสุ่มตัวอย่าง ซึ่งส่งผลต่อความเอนเอียงในการประมาณค่า โดยที่ความเอนเอียงจะมากขึ้นเมื่อขนาดตัวอย่างเพิ่มขึ้น ดังนั้นหากไม่ควบคุมให้ดีความคลาดเคลื่อนจากกรอบตัวอย่างจะส่งผลต่อการประมาณค่าได้ โดยทางทฤษฎีเราต้องการกรอบตัวอย่างที่สมบูรณ์ เพื่อให้ได้ตัวแทนที่ดีของประชากร เพื่อใช้ค่าที่ได้จากตัวอย่างประมาณค่าพารามิเตอร์ที่สนใจศึกษา โดยมีรายละเอียดดังนี้

1. กรอบตัวอย่างที่สมบูรณ์

กรอบตัวอย่าง คือ บัญชีรายชื่อหรือทะเบียนรายชื่อที่ระบุหน่วยที่มีโอกาสถูกเลือกเป็นหน่วยตัวอย่าง และอาจแสดงที่อยู่ของหน่วยนั้น ๆ ด้วย กรอบตัวอย่างดังกล่าวอาจเชื่อมโยงไปยังทุกหน่วยของประชากรเป้าหมายหรือไม่ก็ได้ โดยขึ้นอยู่กับบัญชีหรือทะเบียนรายชื่อที่มีในขณะที่เราทำการสำรวจ ดังนั้นกรอบตัวอย่างคือสิ่งที่เป็นตัวแทนของประชากรเป้าหมายในลักษณะที่อธิบายประชากรเป้าหมายอย่างเป็นรูปธรรม โดยแสดงทุกหน่วยในประชากรเป้าหมาย เพื่อให้สามารถเลือกหน่วยตัวอย่างได้ รูปแบบของกรอบตัวอย่างอาจมีได้หลากหลาย เช่น ถ้าประชากรเป้าหมายคือครัวเรือนในกรุงเทพฯ กรอบตัวอย่างอาจเป็นรายชื่อของครัวเรือนต่าง ๆ ทุกครัวเรือนที่อยู่ในเขตกรุงเทพฯ หรือกรอบตัวอย่างอาจเป็นแผนที่กรุงเทพฯ ที่แสดงที่ตั้งของครัวเรือนทุกครัวเรือน หรืออาจเป็นแผนที่แสดงอาณาเขตการแจงนับและชุมรุมอาคารต่าง ๆ ซึ่งเมื่อถูกเลือกจะต้องสร้างหรือหารายชื่อครัวเรือน หรือแผนที่ที่แสดงที่ตั้งของครัวเรือนต่อไปอีก เป็นต้น

ในทางทฤษฎีกรอบตัวอย่างที่ใช้ควรเป็นกรอบตัวอย่างที่สมบูรณ์คือกรอบที่ใช้เชื่อมโยงไปยังหน่วยตัวอย่างได้ถูกต้อง ครบถ้วน ไม่ซ้ำซ้อน และทันสมัยหรือเป็นปัจจุบัน เมื่อต้องการศึกษาคุณลักษณะของประชากรจะอธิบายคุณลักษณะของประชากรด้วยตัวอย่าง จะประมาณค่าพารามิเตอร์ที่สนใจศึกษาด้วยค่าสถิติของตัวอย่าง เช่น การประมาณค่าเฉลี่ยประชากร μ ของตัวแปรที่ศึกษา ด้วยการประมาณค่าเฉลี่ยด้วยตัวประมาณอย่างง่าย นอกจากนี้กรอบตัวอย่างที่ใช้ในการสำรวจนั้น จะบ่งบอกแผนการสุ่มตัวอย่างและการประมาณค่าที่เหมาะสมด้วย ทั้งนี้ถ้ากรอบตัวอย่างนั้นมีสาระสำคัญอื่น ๆ อาจนำมาช่วยในการประมาณค่าได้ เช่น ตัวประมาณค่าที่ใช้ตัวแปรที่เกี่ยวข้อง (Composite Estimator) ซึ่งตัวประมาณที่ใช้ตัวแปรที่เกี่ยวข้องที่นิยมใช้ คือ ตัวประมาณอัตราส่วน และตัวประมาณการถดถอย โดยมีรายละเอียดของการประมาณค่าเฉลี่ยประชากร ดังนี้

1.1 การประมาณค่าเฉลี่ยประชากรด้วยตัวประมาณอย่างง่าย เป็นการประมาณค่าเฉลี่ยประชากร μ ของตัวแปรที่ศึกษา โดยสุ่มตัวอย่างแบบง่ายไม่แทนที่ขนาด n จากประชากร N หน่วย เพื่อเก็บรวบรวมค่าของตัวแปรที่ศึกษา แล้วประมาณค่าเฉลี่ยประชากรด้วยค่าเฉลี่ยตัวอย่าง ดังทฤษฎี 1 (ทวีศักดิ์ ศิริพรไพบูลย์, 2549, หน้า 48-50) ดังนี้

ทฤษฎี 1 การประมาณค่าเฉลี่ยประชากรด้วยตัวประมาณอย่างง่าย โดยสุ่มตัวอย่างแบบง่ายไม่แทนที่ขนาด n จากประชากร N หน่วย

1. $\hat{\mu}$ เป็นตัวประมาณค่าเฉลี่ยที่ไม่เอนเอียงของค่าเฉลี่ยประชากร μ

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} y_i \tag{1}$$

2. ความแปรปรวนของตัวประมาณค่าเฉลี่ย

$$V(\hat{\mu}) = \left(1 - \frac{n}{N}\right)\frac{S^2}{n} \tag{2}$$

เมื่อ

$$S^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \mu)^2 \text{ และ } \mu = \frac{1}{N}\sum_{i=1}^{N} y_i$$

3. ค่าประมาณความแปรปรวนของตัวประมาณค่าเฉลี่ย

$$v(\hat{\mu}) = \left(1 - \frac{n}{N}\right)\frac{s^2}{n} \tag{3}$$

เมื่อ

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \hat{\mu})^2$$

ตัวประมาณที่ใช้ตัวแปรช่วยคือ ตัวประมาณอัตราส่วน และตัวประมาณการถดถอย จะมีประสิทธิภาพสูงกว่าตัวประมาณอย่างง่าย เนื่องจากตัวประมาณที่ใช้ตัวแปรช่วยจะทำให้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยน้อยกว่าตัวประมาณที่ไม่ใช้ตัวแปรช่วย เพราะการใช้ตัวแปรช่วยในการประมาณค่าจะทำให้มีสาระเพิ่มมากยิ่งขึ้น โดยเฉพาะตัวแปรช่วยที่มีความสัมพันธ์กับตัวแปรที่ศึกษา และการประมาณค่าจะมีประสิทธิภาพสูงขึ้นเมื่อระดับความสัมพันธ์ของตัวแปรที่ศึกษาและตัวแปรช่วยสูง (สุรินทร์ นิยมมางกูร, 2541) โดยมีรายละเอียดดังนี้

1.2 การประมาณค่าเฉลี่ยประชากรด้วยตัวประมาณอัตราส่วน เป็นการประมาณค่าเฉลี่ยประชากร μ ของตัวแปรที่ศึกษา (Y) โดยใช้ประโยชน์จากตัวแปรช่วย (X) ที่มีความสัมพันธ์กับตัวแปรที่ศึกษา ดังทฤษฎี 2 (ทวีศักดิ์ ศิริพรไพบูลย์, 2549, หน้า 74-78)

ทฤษฎี 2 การประมาณค่าเฉลี่ยประชากรด้วยตัวประมาณอัตราส่วน โดยสุ่มตัวอย่างแบบง่ายไม่แทนที่ขนาด n จากประชากร N หน่วย เมื่อตัวแปรที่ศึกษาและตัวแปรช่วยมีความสัมพันธ์กัน

1. $\hat{\mu}_R$ เป็นตัวประมาณค่าเฉลี่ยที่เอนเอียงของค่าเฉลี่ยประชากร μ

$$\hat{\mu}_R = \hat{R}\mu_x \tag{4}$$

เมื่อ

$$\mu_x = \frac{1}{N}\sum_{i=1}^{N} x_i \text{ และ } \hat{R} = \frac{\hat{\mu}}{\hat{\mu}_x}$$

โดยที่

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} y_i \text{ และ } \hat{\mu}_x = \frac{1}{n}\sum_{i=1}^{n} x_i$$

2. ความแปรปรวนของตัวประมาณค่าเฉลี่ย

$$V(\hat{\mu}_R) = \left(1 - \frac{n}{N}\right)\frac{S_d^2}{n} \tag{5}$$

เมื่อ

$$S_d^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - Rx_i)^2 \text{ และ } R = \frac{\mu}{\mu_x}$$

3. ค่าประมาณความแปรปรวนของตัวประมาณค่าเฉลี่ย

$$v(\hat{\mu}_R) = \left(1 - \frac{n}{N}\right)\frac{s_d^2}{n} \tag{6}$$

เมื่อ

$$s_d^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \hat{R}x_i)^2$$

1.3 การประมาณค่าเฉลี่ยประชากรด้วยตัวประมาณการถดถอย ตัวประมาณการถดถอยเป็นตัวประมาณที่ใช้ประโยชน์จากตัวแปรช่วยในการประมาณค่าเฉลี่ยประชากร μ ของตัวแปรที่ศึกษาเช่นเดียวกับตัวประมาณอัตราส่วน ดังทฤษฎี 3 (ทวีศักดิ์ ศิริพรไพบูลย์, 2549, หน้า 81-84)

ทฤษฎี 3 การประมาณค่าเฉลี่ยประชากรด้วยตัวประมาณการถดถอย โดยสุ่มตัวอย่างแบบง่ายไม่แทนที่ขนาด n จากประชากร N หน่วย เมื่อตัวแปรที่ศึกษา และตัวแปรช่วยมีความสัมพันธ์กันเชิงเส้นตรง โดยที่ $\hat{\beta}$ เป็นค่าประมาณสัมประสิทธิ์ถดถอยด้วยวิธีกำลังสองน้อยที่สุด

1. $\hat{\mu}_L$ เป็นตัวประมาณค่าเฉลี่ยที่เอนเอียงของค่าเฉลี่ยประชากร μ

$$\hat{\mu}_L = \hat{\mu} + \hat{\beta}(\mu_x - \hat{\mu}_x) \qquad (7)$$

เมื่อ

$$\hat{\beta} = \frac{\sum_{i=1}^{n}(x_i - \hat{\mu}_x)(y_i - \hat{\mu})}{\sum_{i=1}^{n}(x_i - \hat{\mu}_x)^2}$$

2. ความแปรปรวนของตัวประมาณค่าเฉลี่ย

$$V(\hat{\mu}_L) = \left(1 - \frac{n}{N}\right)\frac{S_z^2}{n} \qquad (8)$$

เมื่อ

$$S_z^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left[(y_i - \mu) - \beta(x_i - \mu_x)\right]^2$$

โดยที่

$$\beta = \frac{\sum_{i=1}^{N}(x_i - \mu_x)(y_i - \mu)}{\sum_{i=1}^{N}(x_i - \mu_x)^2}$$

3. ค่าประมาณความแปรปรวนของตัวประมาณค่าเฉลี่ย

$$v(\hat{\mu}_L) \approx \left(1 - \frac{n}{N}\right)\frac{s_z^2}{n} \tag{9}$$

เมื่อ

$$s_z^2 = \frac{1}{n-2}\left(\sum_{i=1}^{n}(y_i - \hat{\mu})^2 - \hat{\beta}^2\sum_{i=1}^{n}(x_i - \hat{\mu}_x)^2\right)$$

2. กรอบตัวอย่างไม่สมบูรณ์

กรอบตัวอย่างไม่สมบูรณ์ แบ่งออกเป็น 3 ลักษณะ คือ 1) กรอบตัวอย่างมีการนับซ้ำ (Duplication) เกิดจากการใช้สาระสนเทศมากกว่า 1 แหล่งในการสร้างกรอบตัวอย่าง ทำให้บางหน่วยของประชากรมีการนับซ้ำ ซึ่งหน่วยที่มีการนับซ้ำจะมีโอกาสถูกเลือกเป็นตัวอย่างมากกว่า 1 ครั้ง 2) กรอบตัวอย่างครอบคลุมมากเกินไปเกิดจากการใช้กรอบตัวอย่างที่ครอบคลุมหน่วยที่ไม่ใช่ประชากรเป้าหมาย ทำให้หน่วยตัวอย่างเหล่านี้ถูกเลือกทั้ง ๆ ที่ไม่มีสิทธิ์ และ 3) กรอบตัวอย่างไม่ครอบคลุม เกิดจากการใช้กรอบตัวอย่างที่ไม่ครอบคลุมทุกหน่วยประชากรเป้าหมาย เมื่อทำการเลือกตัวอย่างจากกรอบตัวอย่างก็จะทำให้หน่วยเหล่านี้ไม่มีโอกาสถูกเลือกเลย (สุชาดา กีรนันทน์, 2542, หน้า 366-368)

ปัญหากรอบตัวอย่างไม่สมบูรณ์ในแต่ละกรณีข้างต้น อาจเกิดขึ้นร่วมกันในกรอบตัวอย่างเดียวกัน หรือเกิดขึ้นเพียงอย่างเดียว ซึ่งในกรณีที่เกิดขึ้นร่วมกัน การพิจารณาแก้ไขปัญหาหรือผลที่เกิดขึ้นจะซับซ้อนหรือยุ่งยากมากขึ้น อย่างไรก็ตามปัญหาเหล่านี้มีโอกาสเกิดขึ้นในทางปฏิบัติไม่ว่าจะเกิดปัญหาเดียวหรือมากกว่าหนึ่งปัญหา ทั้งนี้เนื่องจากการสร้างกรอบตัวอย่างเป็นกระบวนการที่ยุ่งยาก ต้องใช้งบประมาณสูง และใช้เวลามากในการจัดสร้าง โดยเฉพาะงานสำรวจขนาดใหญ่ บ่อยครั้งที่ผู้สำรวจไม่สามารถสร้างกรอบตัวอย่างในเรื่องที่ตนเองสนใจศึกษาได้ จึงทำให้ต้องพิจารณากรอบตัวอย่างที่มีอยู่หรือมีผู้อื่นสร้างขึ้น เพื่อนำมาใช้ในการสำรวจ โดยอาจจะทำการปรับปรุงให้เหมาะสมขึ้น หรืออาจไม่สามารถปรับปรุงได้ นอกจากนี้กรอบตัวอย่างที่สมบูรณ์ ณ คาบเวลาหนึ่ง อาจกลายเป็นกรอบตัวอย่างไม่สมบูรณ์ได้เมื่อเวลาผ่านไประยะหนึ่ง ตัวอย่างเช่น การสำรวจครัวเรือน หากช่วงเวลาที่สำรวจนั้นมีครัวเรือนสร้างที่อยู่อาศัยขึ้นใหม่ ครัวเรือนที่สร้างขึ้นใหม่นี้จะไม่มีโอกาสถูกเลือกเป็นตัวอย่าง ซึ่งหากดูจากกรอบตัวอย่างแล้วจะไม่ทราบว่ามีหน่วยประชากรบางหน่วยไม่ถูกรวมอยู่ (Non-included) ในกรอบตัวอย่าง หรือมีหน่วยประชากรบางหน่วยที่ปรากฏรายชื่ออยู่ในกรอบตัวอย่างแต่ไม่ใช่ประชากรเป้าหมาย ณ เวลาที่สำรวจ ดังตัวอย่างทะเบียนธุรกิจ (Business Register: BR) เกี่ยวกับจำนวนของบริษัทที่มีลูกจ้างตั้งแต่ 10 คนขึ้นไปในเดือนพฤศจิกายน พ.ศ. 2543 กับเดือนพฤศจิกายน พ.ศ. 2544 ของประเทศสวีเดน (Ängsved, 2004, p. 6) สมมติว่าในการศึกษาเกี่ยวกับค่าใช้จ่ายในการจัดสวัสดิการแก่ลูกจ้างของบริษัทที่มีลูกจ้างตั้งแต่ 10 คนขึ้นไปในปี พ.ศ. 2544 โดยใช้ BR ในเดือนพฤศจิกายน พ.ศ. 2543 เป็นกรอบตัวอย่างในการเก็บรวบรวมข้อมูล (Reference Period) แต่เนื่องจากมีการวางแผนการสำรวจโดยมีช่วงเวลาเก็บรวบรวมข้อมูลเดือนเมษายนถึงเดือนมิถุนายน พ.ศ. 2544 (Survey Period) ซึ่งภายหลังการสำรวจเมื่อนำกรอบตัวอย่างที่ใช้ในการสำรวจมาเปรียบเทียบกับ BR ในเดือนพฤศจิกายน พ.ศ. 2544 พบว่ากรอบตัวอย่างที่ใช้ในการสุ่มตัวอย่าง เกิดปัญหา ดังนี้ 1) กรอบตัวอย่างไม่ครอบคลุม ซึ่งเกิดจากช่วงเดือนพฤศจิกายน พ.ศ. 2543 ถึงเดือนพฤศจิกายน พ.ศ. 2544 นั้นมีบริษัทที่ก่อตั้งขึ้นใหม่ และมีลูกจ้างตั้งแต่ 10 คนขึ้นไป จำนวน 704 บริษัท และบริษัทที่มีจำนวนลูกจ้างเพิ่มขึ้นจนกระทั่งมีลูกจ้างตั้งแต่ 10 คนขึ้นไปในเดือนพฤศจิกายน พ.ศ. 2544 จำนวน 4,700 บริษัท (ดังตัวเอียงที่แสดงในตาราง 1) ดังนั้นบริษัทที่ก่อตั้งใหม่ และบริษัทที่มีจำนวนพนักงานเพิ่มขึ้นดังกล่าวเป็นประชากรที่เราสนใจศึกษา และ 2) กรอบตัวอย่างครอบคลุมมากเกินไป (ดังตัวหนาที่แสดงในตาราง 1) คือ มีบางบริษัทที่เคยมีลูกจ้างตั้งแต่ 10 คนขึ้นไปในปี พ.ศ. 2543 แต่ในปี พ.ศ. 2544 มีจำนวนลูกจ้างลดลงน้อยกว่า 10 คน จำนวน 3,011 บริษัท และมี

บริษัทที่ปิดกิจการจำนวน 1,317 บริษัท ทำให้บริษัทเหล่านี้ไม่ใช่ประชากรเป้าหมายแต่มีรายชื่ออยู่ในกรอบตัวอย่าง

ตาราง 1 แสดงการเปรียบเทียบทะเบียนธุรกิจ (BR) ของประเทศสวีเดนเกี่ยวกับจำนวนบริษัทระหว่างเดือนพฤศจิกายน พ.ศ. 2543 กับเดือนพฤศจิกายน พ.ศ. 2544

		พฤศจิกายน พ.ศ. 2544			
	จำนวนลูกจ้าง	**น้อยกว่า 10 คน**	**10 คนขึ้นไป**	**ปิดบริษัท**	**รวม**
พฤศจิกายน พ.ศ. 2543	น้อยกว่า 10 คน	720,954	*4,700*	55,195	780,849
	10 คนขึ้นไป	**3,011**	29,069	**1,317**	33,397
	ก่อตั้งใหม่	70,349	*704*	-	71,053
	รวม	794,314	34,473	56,512	

ที่มา: Ängsved. (2004). Estimating the finite population total under frame imperfections. Sweden: Department of Statistics (ESI), Örebro University.

ในบางกรณีที่ไม่สามารถหากรอบตัวอย่างในเรื่องที่สนใจได้ อาจต้องใช้กรอบตัวอย่างสำหรับประชากรอื่นที่ใกล้เคียงกันเพื่อแก้ไขปัญหาการไม่มีกรอบตัวอย่าง โดยยอมรับปัญหาที่เกิดขึ้นตามมา ดังนั้นเมื่อตัดสินใจเลือกกรอบตัวอย่างได้แล้ว ก็ต้องพิจารณาว่ากรอบตัวอย่างมีปัญหาใดหรือไม่ ถ้าพบว่ามีปัญหาก็อาจพยายามแก้ไขกรอบตัวอย่างก่อนเลือกตัวอย่าง เช่น โดยการพิจารณาจากวิธีการสร้างกรอบตัวอย่างว่าได้มาอย่างไร จะปรับปรุงให้ทันสมัยหรือแก้ไขได้อย่างไรบ้าง ถ้าความบกพร่องนั้นมีเพียงเล็กน้อย และการปรับปรุงให้ถูกต้องสมบูรณ์ต้องใช้งบประมาณหรือทรัพยากรมากเกินไป ผู้สำรวจอาจใช้กรอบตัวอย่างนั้นโดยไม่มีการแก้ไข ซึ่งต้องมีการรายงานข้อบกพร่องด้วย เพื่อให้ผู้ใช้ข้อมูลมีความระมัดระวังในการนำผลการสำรวจไปใช้ สำหรับบางสถานการณ์ อาจมีความเป็นไปได้ที่จะปรับประชากรเป้าหมายให้สอดคล้องกับกรอบตัวอย่างที่มีโดยไม่กระทบต่อวัตถุประสงค์ของการสำรวจตัวอย่างมากนัก

การสำรวจด้วยตัวอย่างต้องอาศัยกรอบตัวอย่าง เพื่อใช้เชื่อมโยงไปยังหน่วยตัวอย่างที่ต้องการศึกษา ถ้าใช้กรอบตัวอย่างที่สมบูรณ์ในการสุ่มตัวอย่างก็จะนำมาซึ่งตัวแทนที่ดีของประชากรเป้าหมาย แต่การที่จะได้กรอบตัวอย่างที่สมบูรณ์นั้นในทางปฏิบัติเป็นไปได้ยาก เนื่องจากกรอบตัวอย่างที่มีอยู่มักเป็นกรอบตัวอย่างที่ไม่ทันสมัยหรือไม่เป็นปัจจุบัน ทำให้ในระหว่างช่วงที่มี

การสร้างกรอบตัวอย่างและระหว่างการสำรวจนั้นอาจเว้นระยะเวลานาน ทำให้มีหน่วยประชากรที่สนใจศึกษาบางหน่วยเกิดขึ้นใหม่ และมีบางหน่วยประชากรที่เคยอยู่ในกรอบตัวอย่างได้สูญหายไปหรือไม่ใช่ประชากรที่สนใจศึกษาในขณะที่มีการสำรวจ ในทางปฏิบัติโดยทั่วไปจะไม่ทราบว่ากรอบตัวอย่างที่ใช้นั้นสมบูรณ์หรือไม่ จะทราบก็ต่อเมื่อมีการสัมภาษณ์หรือมีการแจงนับ ดังเช่น Andy, et al. (2008) ได้ศึกษาความเคลื่อนจากการสำรวจทางโทรศัพท์ เนื่องจากปัญหากรอบตัวอย่างไม่สมบูรณ์ และการไม่ได้รับความร่วมมือ พบว่า ก่อนการสำรวจจะไม่ทราบว่าเกิดปัญหากรอบตัวอย่างไม่สมบูรณ์หรือปัญหาการไม่ได้รับความร่วมมือ จนกระทั่งภายหลังการสัมภาษณ์ทางโทรศัพท์ จึงจะสามารถจำแนกได้ว่าหน่วยตัวอย่างใดอยู่หรือไม่อยู่ในกรอบตัวอย่าง และหน่วยตัวอย่างใดให้ความร่วมมือหรือไม่ให้ความร่วมมือ โดยใช้เทคนิคการแบ่งชั้นภูมิภายหลังการแจงนับ ผลการศึกษาพบว่า การสำรวจทางโทรศัพท์มักจะเจอปัญหาการไม่ได้รับความร่วมมือมากกว่าปัญหากรอบตัวอย่างไม่สมบูรณ์

การวิจัยครั้งนี้จะศึกษากรอบตัวอย่างไม่ครอบคลุม ซึ่งมีหน่วยประชากรที่ไม่อยู่ในกรอบตัวอย่าง ทำให้หน่วยประชากรที่ไม่อยู่ในกรอบตัวอย่างไม่มีโอกาสถูกเลือกเป็นตัวอย่าง จึงไม่ได้ตัวแทนของประชากรที่ไม่อยู่ในกรอบตัวอย่าง และถ้าประชากรที่อยู่ในกรอบตัวอย่าง และประชากรที่ไม่อยู่ในกรอบตัวอย่างมีคุณลักษณะต่างกัน ย่อมทำให้ค่าประมาณที่ได้เอนเอียงไปจากค่าที่แท้จริงของประชากร หรือมีความคลาดเคลื่อนสูง ดังนั้นจำเป็นต้องหาแนวทางเพื่อลดความเอนเอียงหรือลดความคลาดเคลื่อนให้น้อยลง

3. **กรอบตัวอย่างไม่ครอบคลุม**

กรอบตัวอย่างไม่ครอบคลุม เกิดขึ้นในกรณีที่ประชากรเป้าหมายมีขนาดใหญ่กว่ากรอบตัวอย่าง และหน่วยของประชากรเป้าหมายซึ่งไม่ปรากฏในกรอบตัวอย่างจะไม่ได้รับโอกาสถูกเลือก แนวทางแก้ไขปัญหานี้เพื่อให้ทุกหน่วยในประชากรเป้าหมายมีโอกาสถูกเลือกเท่าเทียมกันก็คือทำให้หน่วยตัวอย่างเหล่านี้มีโอกาสถูกเลือก ข้อสังเกตประการหนึ่งคือ หน่วยที่ขาดหายไปจากกรอบตัวอย่างนี้ อาจจะกระจุกตัวเป็นกลุ่มหรืออาจจะกระจายอยู่ทั่วไปในประชากรก็ได้ แนวทางแก้ไขอาจทำได้ 2 วิธี (สุชาดา กีรนันทน์, 2542, หน้า 368-370) คือ

3.1 การใช้กรอบตัวอย่างเสริม (Supplementary Sampling Frames) วิธีนี้เหมาะสมกับกรณีที่หน่วยที่ขาดหายไปจากกรอบตัวอย่างรวมกันเป็นกลุ่ม และสามารถหากรอบตัวอย่างอื่นมาเสริมกรอบตัวอย่างเดิม กรอบตัวอย่างที่นำมาเสริมอาจมีเพียงกรอบเดียวหรือมากกว่าหนึ่งกรอบก็ได้ กรณีที่ดีที่สุดคือการหากรอบตัวอย่างมาเสริมเพียงหน่วยที่ขาดหายไปเท่านั้น แล้วทำให้กรอบตัวอย่างที่เสริมนี้ครอบคลุมประชากรเป้าหมายพอดี แต่ในทางปฏิบัติการใช้กรอบตัวอย่างเสริมนี้มักจะมีปัญหาที่มีหน่วยตัวอย่างที่ไม่ใช้ประชากรเป้าหมายปรากฏอยู่ หรือมีการซ้ำซ้อนของ

หน่วยตัวอย่าง เช่น การสำรวจตัวอย่างสถานประกอบธุรกิจในกรุงเทพฯ สมมติว่ากรอบตัวอย่างหลักคือ รายชื่อสถานประกอบการต่าง ๆ ที่ได้จากกระทรวงพาณิชย์ แต่เนื่องจากสถานประกอบการขนาดเล็กบางแห่งไม่ได้จดทะเบียนตั้งกิจการ แต่มีกิจกรรมเชิงพาณิชย์ที่น่าศึกษา ดังนั้นหากใช้เพียงกรอบตัวอย่างเดียวจะมีปัญหาคือกรอบตัวอย่างไม่ครอบคลุม จึงมีผู้เสนอให้ใช้กรอบตัวอย่างเสริม เช่น แผนที่แสดงเขตกรุงเทพ เพื่อนำสถานที่ประกอบการขนาดเล็กที่ไม่ได้จดทะเบียนเข้ามาด้วย ในลักษณะการเลือกตัวอย่างเมื่อหน่วยตัวอย่างเป็นพื้นที่ย่อย ซึ่งวิธีการนี้แม้จะแก้ปัญหาการขาดตกสถานประกอบการขนาดเล็ก แต่ก็อาจต้องแก้ปัญหาความซ้ำซ้อนอีก เนื่องจากสถานประกอบการอาจปรากฏอยู่ทั้งกรอบตัวอย่างหลักและกรอบตัวอย่างเสริม อีกทั้งการใช้กรอบตัวอย่างเสริมจะต้องเป็นกรณีที่แบ่งกลุ่มหรือประเภทที่ชัดเจน ทำให้บางครั้งในทางปฏิบัติไม่สามารถหากรอบตัวอย่างเสริมได้

3.2 การเพิ่มหน่วยที่ขาดหายด้วยการผนวกหน่วยประชากรเป้าหมายที่ไม่อยู่ในกรอบตัวอย่างเข้ามา (Half-open Interval) ในกรณีที่หน่วยที่ขาดหายไปในประชากร เช่น การสำรวจตัวอย่างเกี่ยวกับครู สมมติว่ากรอบตัวอย่างคือ บัญชีรายชื่อครูของกรมสามัญศึกษา แต่เนื่องจากบัญชีรายชื่อนี้เป็นรายชื่อตามบัญชีเงินเดือนที่แจ้งไว้ตามปีงบประมาณ จึงขาดความทันสมัยอยู่บ้าง เช่น มีครูที่เพิ่งรับตำแหน่ง และอยู่ในกระบวนการดำเนินงานเรื่องเงินเดือน ทำให้บัญชีรายชื่อที่นำมาใช้เป็นกรอบตัวอย่างมีปัญหาขาดตกชื่อของครูใหม่ของแต่ละโรงเรียนไป หรือในกรณีที่ใช้แผนที่ตั้งของอาคารบ้านเรือนเป็นกรอบตัวอย่าง ถ้าแผนที่จัดทำขึ้นมาระยะเวลาหนึ่งแล้ว อาจมีครัวเรือนสร้างครัวเรือนใหม่ในแต่ละพื้นที่ซึ่งไม่ปรากฏในกรอบตัวอย่าง วิธีการผนวกหน่วยประชากรเป้าหมายที่ไม่อยู่ในกรอบตัวอย่างเข้ามาในกรอบตัวอย่าง หรือวิธี **Half-open Interval** เป็นวิธีที่เสนอโดย Hensen, Hurwitz and Madow ในปี ค.ศ. 1953 มีหลักการดังนี้ เมื่อเรียงหน่วยต่าง ๆ ในกรอบตัวอย่างแล้วให้ถือว่าหน่วยประชากรที่ไม่อยู่ในกรอบตัวอย่างที่ขาดหายไปจากกรอบตัวอย่างระหว่างหน่วยคู่ใดในกรอบตัวอย่างเป็นส่วนผนวกของหน่วยประชากรที่อยู่ในกรอบตัวอย่างก่อนหน้านั้น ดังนั้นแม้ว่ากรอบตัวอย่างที่ใช้จะมีประชากรเป้าหมายหน่วยใดที่ขาดตกไปจากกรอบตัวอย่าง แต่ภายใต้แนวคิดนี้ ย่อมถือได้ว่าทุกหน่วยจะถูกครอบคลุมด้วยกรอบตัวอย่างในลักษณะปรากฏอยู่ หรือผนวกเข้ากับหน่วยประชากรที่อยู่ในกรอบตัวอย่าง นั่นคือ ในทางปฏิบัติเมื่อเลือกหน่วยตัวอย่าง และเก็บรวบรวมข้อมูลได้แล้วจะทำการตรวจสอบว่า หน่วยที่พบในภาคสนามที่อยู่ถัดไปจากหน่วยตัวอย่างที่ได้ปรากฏนั้นอยู่ในกรอบตัวอย่างหรือไม่ ถ้าปรากฏในกรอบตัวอย่างแล้ว ก็จะดำเนินการเก็บรวบรวมข้อมูลจากหน่วยตัวอย่างถัดไป แต่ถ้าหน่วยนั้นไม่ปรากฏในกรอบตัวอย่าง ผู้สำรวจต้องทำการเก็บรวบรวมข้อมูลจากหน่วยนั้น และตรวจสอบต่อไป

อีกว่าหน่วยที่อยู่ถัดออกไปอีกปรากฏอยู่ในกรอบตัวอย่างหรือไม่ ถ้าไม่ก็ต้องทำการการเก็บข้อมูลอีกจนกระทั่งพบหน่วยที่ปรากฏในกรอบตัวอย่างจึงจะหยุดกระบวนการนี้ จะเห็นว่าด้วยวิธีการนี้หน่วยตัวอย่างทุกหน่วยจะมีโอกาสถูกเลือกด้วยสภาพตัวของมันเองหรือผ่านหน่วยอื่น กล่าวคือหน่วยประชากรที่ไม่อยู่ในกรอบตัวอย่างจะมีโอกาสถูกเลือกเท่ากับหน่วยประชากรที่อยู่ในกรอบตัวอย่าง ดังนั้นถ้าการเลือกตัวอย่างเป็นการเลือกตัวอย่างแบบง่ายขนาด n_1 จากประชากรในกรอบตัวอย่าง N_1 หน่วย ดังนั้นประชากรที่ไม่อยู่ในกรอบตัวอย่างก็จะมีโอกาสถูกเลือก n_1 / N_1 ด้วย ซึ่งวิธี Half-open Interval นับได้ว่าเป็นวิธีการที่สะดวก หากการตรวจสอบว่าหน่วยที่อยู่ถัดไปปรากฏในกรอบตัวอย่างหรือไม่นั้น ทำได้ง่ายพอสมควร การกำหนดการเรียงลำดับของหน่วยต่าง ๆ หรือลักษณะของหน่วยต่าง ๆ ปรากฏในประชากรจึงเป็นสิ่งสำคัญ เพื่อจะได้จัดเรียงหน่วยต่าง ๆ ที่เป็นหน่วยผนวกได้โดยไม่เกิดปัญหา นอกจากนี้ขนาดตัวอย่างจะเพิ่มสูงขึ้น ซึ่งจะทำให้ค่าใช้จ่ายในการเก็บรวบรวมข้อมูลและการประมวลผลเพิ่มสูงขึ้นได้

ในทางปฏิบัติโดยทั่วไปจะไม่ทราบว่ากรอบตัวอย่างที่ใช้นั้นเป็นกรอบตัวอย่างไม่ครอบคลุมหรือไม่ แต่ถ้ามีการคาดการณ์ว่ากรอบตัวอย่างที่ใช้ในการสุ่มตัวอย่างนั้นเป็นกรอบตัวอย่างไม่ครอบคลุม เนื่องจากในช่วงเวลาระหว่างการจัดทำกรอบตัวอย่างจนกระทั่งถึงช่วงที่มีการสำรวจนั้นเว้นช่วงระยะเวลานาน หรือการสร้างกรอบตัวอย่างนั้นมีขั้นตอนในการจัดทำที่ไม่น่าเชื่อถือ เราจะคาดหวังว่าจะสามารถแบ่งประชากรออกเป็น 2 ชั้นภูมิภายหลังการแจงนับ (Poststratum) คือ หน่วยประชากรที่อยู่ในกรอบตัวอย่าง (Included Units) และหน่วยประชากรที่ไม่อยู่ในกรอบตัวอย่าง (Non-included Units) ดังนั้นในระหว่างการสำรวจตัวอย่างจากกรอบตัวอย่างนั้น จะทำการเก็บรวบรวมตัวอย่างที่ไม่อยู่ในกรอบตัวอย่างที่เป็นประชากรที่สนใจศึกษาด้วย โดยตรวจสอบกับรายชื่อหรือที่อยู่ที่ปรากฏอยู่ในกรอบตัวอย่างหรือไม่ ถ้าหน่วยตัวอย่างนั้นไม่ปรากฏในกรอบตัวอย่าง แสดงว่าหน่วยตัวอย่างที่ได้นั้นจะเป็นหน่วยประชากรที่ไม่อยู่ในกรอบตัวอย่าง

การสุ่มตัวอย่างจากกรอบตัวอย่างไม่ครอบคลุม จะพิจารณาให้ 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 จะพบว่าไม่มีกรอบตัวอย่างสำหรับประชากร U_2 และเกิดปัญหาในการวิเคราะห์ข้อมูล เช่น การประมาณค่าเฉลี่ยประชากรของตัวแปรที่สนใจศึกษา 3 ประเด็น คือ 1) ไม่มีตัวอย่างสำหรับประชากรที่ไม่อยู่ในกรอบตัวอย่าง 2) ไม่ทราบค่า N_2 และ 3)

ไม่ทราบสาระสนเทศของประชากร U_2 เช่น ไม่ทราบค่าเฉลี่ยประชากรของตัวแปรช่วย μ_{x2} โดยที่ตัวแปรช่วยจะมีความสัมพันธ์เชิงเส้นตรงกับตัวแปรที่ศึกษา ซึ่งการประมาณค่าที่ใช้ตัวแปรช่วยจะให้การประมาณค่าที่แม่นยำ เมื่อตัวแปรที่ศึกษาและตัวแปรช่วยมีความสัมพันธ์กันสูง ดังนั้นจากปัญหา

1. ไม่มีตัวอย่างสำหรับประชากรที่ไม่อยู่ในกรอบตัวอย่าง จะแก้ไขปัญหาโดยการสุ่มตัวอย่างด้วยวิธี **Half-open Interval** ที่เสนอโดย Hensen, Hurwitz and Madow

2. ไม่ทราบค่า N_2 จากตัวประมาณค่า N_2 ที่เสนอโดย Agarwal and Gupta (2008) จากการสุ่มตัวอย่างแบบง่ายไม่แทนที่ (Simple Random Sampling Without Replacement: SRSWOR) เมื่อเกิดปัญหากรอบตัวอย่างไม่ครอบคลุม

3. ไม่ทราบค่าเฉลี่ยประชากรของตัวแปรช่วย μ_{x2} เนื่องจากไม่ทราบ μ_{x2} เมื่อ μ_{x1} และ μ_{x2} สอดคล้องกัน จะประมาณค่า μ_{x2} ด้วยค่าเฉลี่ยตัวแปรช่วย μ_{x1} ของประชากรที่อยู่ในกรอบตัวอย่าง ดังงานวิจัยของ Ängsved (2004) ที่ศึกษาการประมาณค่าผลรวมประชากรด้วยตัวประมาณการถดถอย ภายใต้กรอบตัวอย่างไม่สมบูรณ์ โดยจะนำเสนอการสุ่มตัวอย่างแบบง่ายไม่แทนที่ และการประมาณค่าเฉลี่ยประชากร เมื่อเกิดปัญหากรอบตัวอย่างไม่ครอบคลุมดังต่อไปนี้

4. การสุ่มตัวอย่างแบบง่ายไม่แทนที่ และการประมาณค่าเฉลี่ยประชากร เมื่อเกิดปัญหากรอบตัวอย่างไม่ครอบคลุม

การสุ่มตัวอย่างแบบง่ายไม่แทนที่ เมื่อเกิดปัญหากรอบตัวอย่างไม่ครอบคลุม จะกระทำได้ก็ต่อเมื่อกรอบตัวอย่างที่ใช้ในการสุ่มตัวอย่างนั้น เป็นกรอบตัวอย่างที่ระบุตำแหน่งที่ตั้งหรือที่อยู่ของหน่วยตัวอย่างที่ชัดเจนและแน่นอน เช่น ครัวเรือน สถานประกอบการ เป็นต้น และในบางครั้งกรอบตัวอย่างที่ใช้นั้นจำเป็นต้องมีแผนที่สำหรับใช้ในการระบุตำแหน่งที่ตั้งของหน่วยตัวอย่างด้วย และหน่วยประชากรจะกระจายอยู่ทั่วไปในประชากร ดังนั้นจึงทำให้สามารถหาที่อยู่หรือหน่วยตัวอย่างที่อยู่ในกรอบตัวอย่าง และหาหน่วยตัวอย่างที่ไม่อยู่ในกรอบตัวอย่างได้

สมมติไม่ทราบว่ากรอบตัวอย่างที่ใช้ในการสำรวจตัวอย่างเป็นกรอบตัวอย่างไม่ครอบคลุมหรือไม่ แต่ภายหลังการสำรวจพบว่า จากการสุ่มตัวอย่างขนาด n_1 จากประชากรในกรอบตัวอย่าง N_1 หน่วย เมื่อมีการตรวจสอบรายชื่อของหน่วยตัวอย่างกับรายชื่อในกรอบตัวอย่างพบว่า มีตัวอย่างที่ไม่อยู่ในกรอบตัวอย่าง n_2 หน่วย และเมื่อพิจารณาพบว่าตัวอย่างขนาด n_2 เป็นหน่วยตัวอย่างที่อยู่ในประชากรที่ต้องการศึกษาทำให้ไม่สามารถละทิ้งได้ ดังนั้นจึงเกิดปัญหากรอบตัวอย่างไม่ครอบคลุม และเราคาดหวังว่าการแยกตัวอย่างออกเป็นชั้นภูมิระหว่างหน่วยตัวอย่างที่อยู่ในกรอบตัวอย่าง และหน่วยตัวอย่างที่ไม่อยู่ในกรอบตัวอย่าง จะช่วยให้การวิเคราะห์ข้อมูลมีคุณภาพดีขึ้น แต่เราไม่สามารถสุ่มตัวอย่างแบบมีชั้นภูมิในขั้นตอนของการสำรวจได้

เนื่องจากไม่ทราบจำนวนประชากรที่ไม่อยู่ในกรอบตัวอย่าง และไม่มีกรอบตัวอย่างสำหรับหน่วยตัวอย่างที่ไม่อยู่ในกรอบตัวอย่าง จึงจำเป็นต้องใช้วิธีการแบ่งชั้นภูมิภายหลังการแจงนับ และจะเห็นว่า n_2 เปลี่ยนแปลงไปตามการสุ่มตัวอย่างในแต่ละครั้ง ถึงแม้จะกำหนดขนาดตัวอย่าง n_1 คงที่ ดังนั้นจะใช้วิธีการแบ่งชั้นภูมิภายหลังการแจงนับร่วมกับวิธี Half-open Interval เพื่อแก้ไขปัญหากรอบตัวอย่างไม่ครอบคลุม

ภายหลังการสำรวจตัวอย่างจากกรอบตัวอย่างไม่ครอบคลุม จะสามารถแบ่งประชากรออกเป็น 2 ชั้นภูมิภายหลังการแจงนับ โดยใช้วิธีการแบ่งชั้นภูมิภายหลังการแจงนับ ซึ่งจะกำหนดให้

ชั้นภูมิภายหลังการแจงนับที่ $h = 1$ คือ หน่วยตัวอย่างที่อยู่ในกรอบตัวอย่าง และ

ชั้นภูมิภายหลังการแจงนับที่ $h = 2$ คือ หน่วยตัวอย่างที่ไม่อยู่ในกรอบตัวอย่าง

Agarwal and Gupta (2008) ศึกษาการสุ่มตัวอย่างแบบง่ายไม่แทนที่เมื่อเกิดปัญหากรอบตัวอย่างไม่ครอบคลุมว่ามีประชากรที่อยู่ในกรอบตัวอย่าง N_1 หน่วย และมีประชากรที่ไม่อยู่ในกรอบตัวอย่าง N_2 หน่วย ในการประมาณค่าเฉลี่ยประชากรที่สนใจศึกษาขนาด N_T เมื่อไม่มีกรอบตัวอย่างสำหรับประชากรที่ไม่อยู่ในกรอบตัวอย่าง จะกำหนดให้

y_{1i} แทน ค่าของตัวแปรที่ศึกษาที่อยู่ในกรอบตัวอย่าง หน่วยที่ i ; $i = 1,\ldots,N_1$

y_{2i} แทน ค่าของตัวแปรที่ศึกษาที่ไม่อยู่ในกรอบตัวอย่าง หน่วยที่ i ; $i = 1,\ldots,N_2$

x_{1i} แทน ค่าของตัวแปรช่วยที่อยู่ในกรอบตัวอย่าง หน่วยที่ i ; $i = 1,\ldots,N_1$

x_{2i} แทน ค่าของตัวแปรช่วยที่ไม่อยู่ในกรอบตัวอย่าง หน่วยที่ i ; $i = 1,\ldots,N_2$

m_{1i} แทน จำนวนหน่วยตัวอย่างที่ไม่อยู่ในกรอบตัวอย่างระหว่างหน่วยที่ i และ $i+1$ ของหน่วยตัวอย่างที่อยู่ในกรอบตัวอย่าง ; $i = 1,\ldots,N_2$

μ_T แทน ค่าเฉลี่ยประชากรของตัวแปรที่ศึกษาขนาด N_T

μ_1 แทน ค่าเฉลี่ยประชากรที่อยู่ในกรอบตัวอย่างของตัวแปรที่ศึกษาขนาด N_1

μ_2 แทน ค่าเฉลี่ยประชากรที่ไม่อยู่ในกรอบตัวอย่างของตัวแปรที่ศึกษาขนาด N_2

μ_{x1} แทน ค่าเฉลี่ยประชากรที่อยู่ในกรอบตัวอย่างของตัวแปรช่วย ขนาด N_1

μ_{x2} แทน ค่าเฉลี่ยประชากรที่ไม่อยู่ในกรอบตัวอย่างของตัวแปรช่วย ขนาด N_2

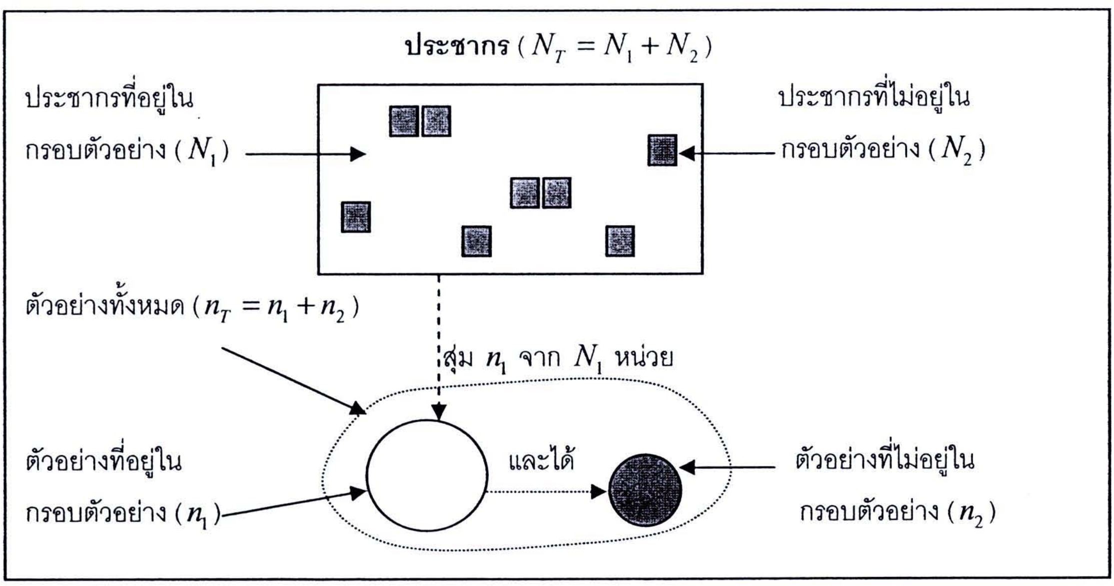
ค่าเฉลี่ยประชากร μ_T เมื่อเกิดปัญหากรอบตัวอย่างไม่ครอบคลุม มีค่าเท่ากับ

$$\mu_T = W_1\mu_1 + W_2\mu_2 \qquad (10)$$

เมื่อ

$$\mu_1 = \frac{1}{N_1}\sum_{i=1}^{N_1} y_{1i} \ , \ \mu_2 = \frac{1}{N_2}\sum_{i=1}^{N_2} y_{2i} \ , \ W_1 = \frac{N_1}{N_T} \text{ และ } W_2 = \frac{N_2}{N_T}$$

การประมาณค่าเฉลี่ยประชากร μ_T เมื่อเกิดปัญหากรอบตัวอย่างไม่ครอบคลุม จะสุ่มตัวอย่างขนาด n_1 จากประชากรในกรอบตัวอย่าง N_1 หน่วย ด้วยวิธี Half-open Interval คือในระหว่างการสำรวจตัวอย่างหน่วยที่ i จะเก็บรวบรวมข้อมูลว่ามีหน่วยตัวอย่างใดบ้างที่ไม่อยู่ในกรอบตัวอย่างระหว่างตัวอย่างที่ i และหน่วยที่ $i+1$ พร้อมทั้งเก็บข้อมูลที่สนใจศึกษาในตัวอย่างที่ไม่อยู่ในกรอบตัวอย่างในระหว่างการสุ่มตัวอย่างขนาด n_1 หน่วย ดังนั้นเราจะได้ตัวอย่างที่อยู่ในกรอบตัวอย่าง n_1 หน่วย และได้ตัวอย่างที่ไม่อยู่ในกรอบตัวอย่างขนาด n_2 ซึ่งจะได้ตัวอย่างทั้งหมด $n_T = n_1 + n_2$ หน่วย (ดังแสดงในภาพ 1)



ภาพ 1 แสดงการสุ่มตัวอย่างเมื่อกรอบตัวอย่างไม่ครอบคลุม

4.1 การประมาณค่าจำนวนประชากรที่ไม่อยู่ในกรอบตัวอย่าง เมื่อเกิดปัญหากรอบตัวอย่างไม่ครอบคลุม

เมื่อกรอบตัวอย่างไม่ครอบคลุม จะมีจำนวนประชากรที่อยู่ในกรอบตัวอย่าง N_1 หน่วย และมีจำนวนประชากรที่ไม่อยู่ในกรอบตัวอย่างขนาด N_2 ดังนั้นประชากรที่สนใจศึกษามี

ทั้งหมด $N_T = N_1 + N_2$ หน่วย แต่เนื่องจากเราไม่มีกรอบตัวอย่างสำหรับประชากรที่ไม่อยู่ในกรอบตัวอย่าง เราจึงไม่ทราบจำนวนและค่าเฉลี่ยของตัวแปรช่วยของประชากรที่ไม่อยู่ในกรอบตัวอย่าง ดังนั้นต้องทำการประมาณค่าเกี่ยวกับค่าประชากรที่ไม่อยู่ในกรอบตัวอย่าง ดังนี้

ทฤษฎี 4 สุ่มตัวอย่างแบบง่ายไม่แทนที่ขนาด n_1 จากประชากรในกรอบตัวอย่างไม่ครอบคลุม N_1 หน่วย

1. $\hat{N}_2$ เป็นตัวประมาณไม่เอนเอียงของจำนวนประชากรที่ไม่อยู่ในกรอบตัวอย่าง N_2

$$\hat{N}_2 = N_1 \frac{1}{n_1} \sum_{i=1}^{n_1} m_{1i} \tag{11}$$

2. ความแปรปรวนของตัวประมาณจำนวนประชากรที่ไม่อยู่ในกรอบตัวอย่าง

$$V\left(\hat{N}_2\right) = N_1^2 \left(1 - \frac{n_1}{N_1}\right) \frac{S_m^2}{n_1} \tag{12}$$

เมื่อ

$$S_m^2 = \frac{1}{N_1 - 1} \sum_{i=1}^{N_1} (m_{1i} - \bar{N}_2)^2 \text{ และ } \bar{N}_2 = \frac{1}{N_1} \sum_{i=1}^{N_1} m_{1i}$$

3. ค่าประมาณความแปรปรวนของตัวประมาณจำนวนประชากรที่ไม่อยู่ในกรอบตัวอย่าง

$$v\left(\hat{N}_2\right) = N_1^2 \left(1 - \frac{n_1}{N_1}\right) \frac{s_m^2}{n_1} \tag{13}$$

เมื่อ

$$s_m^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (m_{1i} - \bar{m})^2 \text{ และ } \bar{m} = \frac{1}{n_1} \sum_{i=1}^{n_1} m_{1i}$$

พิสูจน์

1. เนื่องจาก $\hat{N}_2 = N_1 \dfrac{1}{n_1}\sum_{i=1}^{N_1} a_{1i}m_{1i}$ เมื่อ $a_{1i} \sim Ber\left(\dfrac{1}{N_1}\right)$

$$E\left(\hat{N}_2\right) = N_1\left(\frac{1}{n_1}\sum_{i=1}^{N_1} m_{1i} \cdot E\left(a_{1i}\right)\right)$$

$$= \frac{N_1}{n_1}\sum_{i=1}^{N_1} m_{1i}\left(\frac{n_1}{N_1}\right)$$

$$= N_2 \tag{14}$$

ดังนั้น $\hat{N}_2$ เป็นตัวประมาณที่ไม่เอนเอียงของ N_2

2. $$V\left(\hat{N}_2\right) = N_1^2 V\left[\frac{1}{n_1}\sum_{i=1}^{n} m_{1i}\right]$$

$$= N_1^2 V\left[\hat{\bar{N}}_2\right]$$

$$= N_1^2\left(1 - \frac{n_1}{N_1}\right)\frac{S_m^2}{n_1}$$

เมื่อ

$$S_m^2 = \frac{1}{N_1 - 1}\sum_{i=1}^{N_1}(m_{1i} - \bar{N}_2)^2 \text{ และ } \bar{N}_2 = \frac{1}{N_1}\sum_{i=1}^{N_1} m_{1i}$$

3. เนื่องจาก s_m^2 เป็นค่าประมาณ S_m^2 จะได้ค่าประมาณความแปรปรวนของตัวประมาณจำนวนประชากรที่ไม่อยู่ในกรอบตัวอย่าง

$$v\left(\hat{N}_2\right) = N_1^2\left(1-\frac{n_1}{N_1}\right)\frac{s_m^2}{n_1} \text{ เมื่อ } s_m^2 = \frac{1}{n_1-1}\sum_{i=1}^{n_1}(m_{1i}-\bar{m})^2$$

ทฤษฎี 5 สุ่มตัวอย่างแบบง่ายไม่แทนที่ขนาด n_1 จากประชากรในกรอบตัวอย่างไม่ครอบคลุม N_1 หน่วย

1. $\hat{N}_T$ เป็นตัวประมาณที่ไม่เอนเอียงของจำนวนประชากร N_T

$$\hat{N}_T = N_1 + \hat{N}_2 \tag{15}$$

2. ความแปรปรวนของตัวประมาณจำนวนประชากร

$$V\left(\hat{N}_T\right) = N_1^2\left(1-\frac{n_1}{N_1}\right)\frac{S_m^2}{n_1} \tag{16}$$

3. ค่าประมาณความแปรปรวนของตัวประมาณจำนวนประชากร

$$v\left(\hat{N}_T\right) = N_1^2\left(1-\frac{n_1}{N_1}\right)\frac{s_m^2}{n_1} \tag{17}$$

พิสูจน์

1. จากสมการ (14) ในทฤษฎี 4 จะได้ $E\left(\hat{N}_2\right) = N_2$ ดังนั้นจะได้

$$E\left(\hat{N}_T\right) = E\left[N_1 + \hat{N}_2\right]$$

$$= N_1 + N_2$$

$$= N_T \tag{18}$$

2. $V\left(\hat{N}_T\right) = V(N_1 + \hat{N}_2)$

$$= V(N_1) + V(\hat{N}_2) + 2\text{Cov}(N_1, \hat{N}_2)$$

$$= V(\hat{N}_2)$$

เมื่อ

$$V(N_1) = 0 \text{ และ } \text{Cov}(N_1, \hat{N}_2) = 0$$

และจากสมการ (15) ในทฤษฎี 4 จะได้

$$V\left(\hat{N}_T\right) = N_1^2\left(1 - \frac{n_1}{N_1}\right)\frac{S_m^2}{n_1}$$

3. เนื่องจาก s_m^2 เป็นค่าประมาณ S_m^2 จะได้ค่าประมาณความแปรปรวนของตัวประมาณจำนวนประชากร เท่ากับ

$$v\left(\hat{N}_T\right) = N_1^2\left(1 - \frac{n_1}{N_1}\right)\frac{s_m^2}{n_1} = v\left(\hat{N}_2\right)$$

4.2 การประมาณค่าเฉลี่ยประชากร เมื่อเกิดปัญหากรอบตัวอย่างไม่ครอบคลุม

เมื่อต้องการประมาณค่าเฉลี่ยประชากรของตัวแปรที่ศึกษา เมื่อเกิดปัญหากรอบตัวอย่างไม่ครอบคลุม และภายหลังการแจงนับนั้นทราบว่ามีจำนวนตัวอย่างที่ไม่อยู่ในกรอบตัวอย่าง n_2 หน่วย แต่ไม่ทราบจำนวนประชากรที่ไม่อยู่ในกรอบตัวอย่าง N_2 และค่าเฉลี่ยตัวแปรช่วยของประชากรที่ไม่อยู่ในกรอบตัวอย่าง μ_{x2} เราจะประมาณ N_2 และ μ_{x2} ด้วย $\hat{N}_2$ และ μ_{x1} ตามลำดับ โดยศึกษาการประมาณค่าเฉลี่ยภายหลังการแจงนับ (Post-stratified Mean Estimators) 3 ตัว คือ ตัวประมาณอย่างง่าย ตัวประมาณอัตราส่วน และตัวประมาณการถดถอย โดยมีรายละเอียดดังนี้

ทฤษฎี 6 การประมาณค่าเฉลี่ยประชากรด้วยตัวประมาณอย่างง่าย เมื่อเกิดปัญหากรอบตัวอย่างไม่ครอบคลุม จะได้ตัวประมาณอย่างง่ายภายหลังการแจงนับ ดังนี้

1. $\hat{\mu}_{pst}$ เป็นตัวประมาณค่าเฉลี่ยที่เอนเอียงแบบไม่มีเงื่อนไขของค่าเฉลี่ยประชากร μ_T

$$\hat{\mu}_{pst} = \hat{W}_1\hat{\mu}_1 + \hat{W}_2\hat{\mu}_2 = \sum_{h=1}^{2}\hat{W}_h\hat{\mu}_h \tag{19}$$

เมื่อ

$$\hat{\mu}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} y_{1i}\ ,\ \hat{\mu}_2 = \frac{1}{n_2}\sum_{i=1}^{n_2} y_{2i}\ ,\ \hat{W}_1 = \frac{N_1}{\hat{N}_T} \text{ และ } \hat{W}_2 = \frac{\hat{N}_2}{\hat{N}_T}$$

2. ความแปรปรวนของตัวประมาณแบบไม่มีเงื่อนไขของค่าเฉลี่ย

$$V\left[\hat{\mu}_{pst}\right] \simeq \left(1-\frac{n_T}{\hat{N}_T}\right)\frac{1}{n_T}\sum_{h=1}^{2}\hat{W}_h S_h^2 + \frac{1}{n_T^2}\sum_{h=1}^{2}\left(1-\hat{W}_h\right)S_h^2$$

$$+\frac{1}{N_T^2}\left[W_1^2\mu_1^2 + \left(W_2-1\right)^2\mu_2^2\right] V\left(\hat{N}_2\right) \tag{20}$$

เมื่อ

$$S_h^2 = \frac{1}{N_h-1}\sum_{i=1}^{N_h}(y_{hi}-\mu_h)^2 \text{ โดยที่ } \mu_h = \frac{1}{N_h}\sum_{i=1}^{N_h} y_{hi}$$

3. ค่าประมาณความแปรปรวนแบบไม่มีเงื่อนไขของตัวประมาณค่าเฉลี่ย

$$v\left[\hat{\mu}_{pst}\right] \simeq \left(1-\frac{n_T}{\hat{N}_T}\right)\frac{1}{n_T}\sum_{h=1}^{2}\hat{W}_h s_h^2 + \frac{1}{n_T^2}\sum_{h=1}^{2}\left(1-\hat{W}_h\right)s_h^2$$

$$+\frac{1}{\hat{N}_T^2}\left[\hat{W}_1^2\hat{\mu}_1^2 + \left(\hat{W}_2-1\right)^2\hat{\mu}_2^2\right] v\left(\hat{N}_2\right) \tag{21}$$

เมื่อ

$$s_h^2 = \frac{1}{n_h-1}\sum_{i=1}^{n_h}(y_{hi}-\hat{\mu}_h)^2$$

พิสูจน์ (ทวีศักดิ์ ศิริพรไพบูลย์, 2549, หน้า 114-115) เนื่องจาก n_2 เป็นตัวแปรที่สามารถเปลี่ยนแปลงได้ตามขนาดตัวอย่าง n_1 ในการสุ่มตัวอย่างแต่ละครั้ง ดังนั้นภายใต้เงื่อนไขที่กำหนดให้ n_1 และ n_2 คงที่ จะได้ตัวประมาณที่เอนเอียงแบบมีเงื่อนไข (Conditional Biased) คือ

$$\begin{aligned} E\left[\hat{\mu}_{pst} \mid n_1, n_2\right] &\simeq \sum_{h=1}^{2} \hat{W}_h\, E\left[\hat{\mu}_h \mid n_1, n_2\right] \\ &= \hat{W}_1\mu_1 + \hat{W}_2\mu_2 \\ &= \frac{N_1}{\hat{N}_T}\mu_1 + \frac{\hat{N}_2}{\hat{N}_T}\mu_2 \\ &\neq \mu_T \end{aligned} \tag{22}$$

และ

$$\begin{aligned} V\left[\hat{\mu}_{pst} \mid n_1, n_2\right] &\simeq \sum_{h=1}^{2} \hat{W}_h\, V\left[\hat{\mu}_h \mid n_1, n_2\right] \\ &= \sum_{h=1}^{2} \hat{W}_h \left(1 - \frac{n_h}{N_h}\right)\frac{S_h^2}{n_h} \end{aligned} \tag{23}$$

และจากข้อเท็จจริงที่สำคัญ 2 ประการ (Cochran, 1977) มีดังนี้

$E(X) = E_1\left[E_2(X \mid Z)\right]$ และ $V(X) = E_1\left[V_2(X \mid Z)\right] + V_1\left[E_2(X \mid Z)\right]$

เมื่อ

E_1 และ V_1 คือ ค่าคาดหวัง และค่าความแปรปรวนของ X เมื่อ Z เป็นตัวแปรสุ่ม

E_2 และ V_2 คือ ค่าคาดหวัง และค่าความแปรปรวนของ X เมื่อ Z เป็นค่าคงที่

จากสมการ (22) จะได้ตัวประมาณที่เอนเอียงแบบไม่มีเงื่อนไข (Unconditional Biased) ดังนี้

$$E\left[\hat{\mu}_{pst}\right] \simeq E_1\left[\sum_{h=1}^{2}\hat{W}_h \cdot E_2\left(\hat{\mu}_h \mid n_1, n_2\right)\right]$$

$$= E_1\left[\hat{W}_1\mu_1 + \hat{W}_2\mu_2\right]$$

$$= \frac{N_1}{\hat{N}_T}\mu_1 + \frac{\hat{N}_2}{\hat{N}_T}\mu_2$$

$$\neq \mu_T \tag{24}$$

และ

$$V\left[\hat{\mu}_{pst}\right] = E_1\left[V_2\left(\hat{\mu}_{pst} \mid n_1, n_2\right)\right] + V_1\left[E_2\left(\hat{\mu}_{pst} \mid n_1, n_2\right)\right] \tag{25}$$

จะพิจารณา

$$E_1\left[V_2\left(\hat{\mu}_{pst} \mid n_1, n_2\right)\right] = E_1\left[\sum_{h=1}^{2}\hat{W}_h^2\left(\frac{1}{n_h} - \frac{1}{N_h}\right)S_h^2\right]$$

$$\simeq \sum_{h=1}^{2}\hat{W}_h^2\left(E_1\left[\frac{1}{n_h}\right] - \frac{1}{N_h}\right)S_h^2$$

จาก Cochran (1977, p. 135 อ้างอิงใน ทวีศักดิ์ ศิริพรไพบูลย์, 2549, หน้า 114) สามารถแสดงได้ว่า

$$E\left[\frac{1}{n_h}\right] \simeq \frac{1}{n_T\hat{W}_h} + \frac{1-\hat{W}_h}{n_T^2\hat{W}_h^2}$$

ดังนั้นจะได้

$$E_1\left[V_2\left(\hat{\mu}_{pst} \mid n_1,n_2\right)\right] = \sum_{h=1}^{2}\hat{W}_h^2 S_h^2\left(\frac{1}{n_T\hat{W}_h}+\frac{1-\hat{W}_h}{n_T^2\hat{W}_h^2}\right)-\sum_{h=1}^{2}\hat{W}_h^2\frac{1}{N_h}S_h^2$$

$$\simeq \frac{1}{n_T}\sum_{h=1}^{2}\hat{W}_h S_h^2 + \frac{1}{n_T^2}\sum_{h=1}^{2}\left(1-\hat{W}_h\right)S_h^2 - \frac{1}{\hat{N}_T}\sum_{h=1}^{2}\hat{W}_h S_h^2$$

$$=\left(\frac{1}{n_T}-\frac{1}{\hat{N}_T}\right)\sum_{h=1}^{2}\hat{W}_h S_h^2 + \frac{1}{n_T^2}\sum_{h=1}^{2}\left(1-\hat{W}_h\right)S_h^2$$

$$=\left(1-\frac{n_T}{\hat{N}_T}\right)\frac{1}{n_T}\sum_{h=1}^{2}\hat{W}_h S_h^2 + \frac{1}{n_T^2}\sum_{h=1}^{2}\left(1-\hat{W}_h\right)S_h^2 \qquad (26)$$

และพิจารณา

$$V_1\left[E_2\left(\hat{\mu}_{pst} \mid n_1,n_2\right)\right] = V_1\left(\frac{N_1}{\hat{N}_T}\mu_1+\frac{\hat{N}_2}{\hat{N}_T}\mu_2\right)$$

$$= V_1\left(\hat{W}_1\mu_1+\hat{W}_2\mu_2\right)$$

$$= \mu_1^2\ V_1\left(\hat{W}_1\right)+\mu_2^2\ V_1\left(\hat{W}_2\right) \qquad (27)$$

สำหรับการพิจารณา $V_1\left(\hat{W}_1\right)$ และ $V_1\left(\hat{W}_2\right)$ จะพิจารณาจากการใช้ฟังก์ชันกำลังที่หนึ่ง (Linearization) ของเทย์เลอร์ (Taylor) หาค่าประมาณความแปรปรวนของ $\hat{W}_1 = N_1/\hat{N}_T$ และ $\hat{W}_2 = \hat{N}_2/\hat{N}_T$ ในการประมาณ $W_1 = N_1/N_T$ และ $W_2 = N_2/N_T$ ของประชากร ตามลำดับ ซึ่งพิสูจน์ทำนองเดียวกับตัวประมาณอัตราส่วน $\hat{R} = \hat{\mu}/\hat{\mu}_x$ (ประชุม สุวัตถี, 2552, หน้า 537-538) โดยกำหนดให้

$$W_1 = g_1(N_1,N_T) = \frac{N_1}{N_T}$$

$$W_2 = g_2(N_2, N_T) = \frac{N_2}{N_T}$$

$$\frac{\partial g_1}{\partial N_1}(N_1, N_T) = \frac{1}{N_T} \text{ และ } \frac{\partial g_1}{\partial N_T}(N_1, N_T) = -\frac{N_1}{N_T^2}$$

$$\frac{\partial g_2}{\partial N_2}(N_2, N_T) = \frac{1}{N_T} \text{ และ } \frac{\partial g_2}{\partial N_T}(N_2, N_T) = -\frac{N_2}{N_T^2}$$

พิจารณา

$$\hat{W}_1 = g_1(N_1, \hat{N}_T) = g_1(N_1, N_T) + \frac{\partial g_1}{\partial N_1}\left(\hat{N}_1 - N_1\right) + \frac{\partial g_1}{\partial N_T}\left(\hat{N}_T - N_T\right)$$

และเนื่องจาก N_1 ทราบค่า จึงไม่จำเป็นจะต้องประมาณค่า N_1 ด้วย $\hat{N}_1$ ดังนั้นจะได้

$$\hat{W}_1 = W_1 + \frac{1}{N_T}(N_1 - N_1) - \frac{N_1}{N_T^2}\left(\hat{N}_T - N_T\right)$$

$$= W_1 - \frac{N_1}{N_T^2}\left(\hat{N}_T - N_T\right)$$

ค่าคาดหวังของ $\hat{W}_1$ มีค่าเท่ากับ

$$E\left(\hat{W}_1\right) = W_1 - \frac{N_1}{N_T^2}\left[E\left(\hat{N}_T\right) - N_T\right]$$

$$= W_1 - \frac{N_1}{N_T^2}\left[N_T - N_T\right]$$

$$= W_1 \qquad (28)$$

และ

$$\hat{W}_1 - W_1 = -\frac{N_1}{N_T^2}\left(\hat{N}_T - N_T\right)$$

จะได้

$$V\left(\hat{W}_1\right) \approx MSE\left(\hat{W}_1\right)$$

$$= E\left(\hat{W}_1 - W_1\right)^2$$

$$= E\left[-\frac{N_1}{N_T^2}\left(\hat{N}_T - N_T\right)\right]^2$$

$$= \frac{N_1^2}{N_T^4}V\left(\hat{N}_T\right)$$

$$= \frac{1}{N_T^2}W_1^2V\left(\hat{N}_T\right)$$

จากทฤษฎี 5 จะได้ $V\left(\hat{N}_T\right) = V\left(\hat{N}_2\right)$ ดังนั้น

$$V\left(\hat{W}_1\right) = \frac{1}{N_T^2}W_1^2V\left(\hat{N}_2\right) \tag{29}$$

และพิจารณา

$$\hat{W}_2 = g_2(\hat{N}_2, \hat{N}_T)$$

$$= g_2(N_2, N_T) + \frac{\partial g_2}{\partial N_2}\left(\hat{N}_2 - N_2\right) + \frac{\partial g_2}{\partial N_T}\left(\hat{N}_T - N_T\right)$$

$$= W_2 + \frac{1}{N_T}\left(\hat{N}_2 - N_2\right) - \frac{N_2}{N_T^2}\left(\hat{N}_T - N_T\right)$$

ค่าคาดหวังของ $\hat{W}_2$ มีค่าเท่ากับ

$$E\left(\hat{W}_2\right) = W_2 + \frac{1}{N_T}\left[E\left(\hat{N}_2\right) - N_2\right] - \frac{N_2}{N_T^2}\left[E\left(\hat{N}_T\right) - N_T\right]$$

$$= W_2 + \frac{1}{N_T}\left[N_2 - N_2\right] - \frac{N_2}{N_T^2}\left[N_T - N_T\right]$$

$$= W_2 \tag{30}$$

และ

$$\hat{W}_2 - W_2 = \frac{1}{N_T}\left(\hat{N}_2 - N_2\right) - \frac{N_2}{N_T^2}\left(\hat{N}_T - N_T\right)$$

จะได้

$$V\left(\hat{W}_2\right) \approx MSE\left(\hat{W}_2\right)$$

$$= E\left(\hat{W}_2 - W_2\right)^2$$

$$= E\left[\frac{1}{N_T}\left(\hat{N}_2 - N_2\right) - \frac{N_2}{N_T^2}\left(\hat{N}_T - N_T\right)\right]^2$$

$$= \frac{1}{N_T^2}\ V\left(\hat{N}_2\right) - 2\frac{N_2}{N_T^3}\ \mathrm{Cov}\left(\hat{N}_2, \hat{N}_T\right) + \frac{N_2^2}{N_T^4}\ V\left(\hat{N}_T\right)$$

$$= \frac{1}{N_T^2}\left[V\left(\hat{N}_2\right) - 2\frac{N_2}{N_T}\ \mathrm{Cov}\left(\hat{N}_2, \hat{N}_T\right) + \frac{N_2^2}{N_T^2}\ V\left(\hat{N}_T\right)\right]$$

$$= \frac{1}{N_T^2}\left[V\left(\hat{N}_2\right) - 2W_2 \ \mathrm{Cov}\left(\hat{N}_2, \hat{N}_T\right) + W_2^2 \ V\left(\hat{N}_T\right)\right]$$

จากทฤษฎี 5 จะได้ $V\left(\hat{N}_T\right) = V\left(\hat{N}_2\right)$ และ $\hat{N}_T = N_1 + \hat{N}_2$ ดังนั้นจะได้

$$V\left(\hat{W}_2\right) = \frac{1}{N_T^2}\left[V\left(\hat{N}_2\right) - 2W_2 \ V\left(\hat{N}_2\right) + W_2^2 \ V\left(\hat{N}_2\right)\right]$$

$$= \frac{1}{N_T^2}\left(W_2^2 - 2W_2 + 1\right)V\left(\hat{N}_2\right)$$

$$= \frac{1}{N_T^2}\left(W_2 - 1\right)^2 V\left(\hat{N}_2\right) \qquad (31)$$

แทนค่าสมการ (29) และ (31) ในสมการ (27) จะได้

$$V_1\left[E_2\left(\hat{\mu}_{pst} \mid n_1, n_2\right)\right] = \mu_1^2 \ \frac{1}{N_T^2} W_1^2 V\left(\hat{N}_2\right) + \mu_2^2 \ \frac{1}{N_T^2}\left(W_2 - 1\right)^2 V\left(\hat{N}_2\right)$$

$$= \frac{1}{N_T^2}\left[W_1^2 \mu_1^2 + \left(W_2 - 1\right)^2 \mu_2^2\right] V\left(\hat{N}_2\right) \qquad (32)$$

แทนค่าสมการ (26) และ (32) ในสมการ (25) จะได้ค่าความแปรปรวนของ $\hat{\mu}_{pst}$ เท่ากับ

$$V\left(\hat{\mu}_{pst}\right) \simeq \left(1 - \frac{n_T}{\hat{N}_T}\right)\frac{1}{n_T}\sum_{h=1}^{2} \hat{W}_h S_h^2 + \frac{1}{n_T^2}\sum_{h=1}^{2}\left(1 - \hat{W}_h\right) S_h^2$$

$$+ \frac{1}{N_T^2}\left[W_1^2 \mu_1^2 + \left(W_2 - 1\right)^2 \mu_2^2\right] V\left(\hat{N}_2\right)$$

เมื่อ

$$V\left(\hat{N}_2\right) = N_1^2\left(1-\frac{n_1}{N_1}\right)\frac{S_m^2}{n_1} \text{ (จากทฤษฎี 4)}$$

ซึ่งอาจประมาณค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของ $\hat{W}_1$ ด้วย

$$v\left(\hat{W}_1\right) = \frac{1}{\hat{N}_T^2}\hat{W}_1^2\, v\left(\hat{N}_2\right) \tag{33}$$

และอาจประมาณค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของ $\hat{W}_2$ ด้วย

$$v\left(\hat{W}_2\right) = \frac{1}{\hat{N}_T^2}\left(\hat{W}_2 - 1\right)^2 v\left(\hat{N}_2\right) \tag{34}$$

จึงสามารถประมาณ $V\left(\hat{W}_1\right)$ ด้วย $v\left(\hat{W}_1\right)$ และประมาณ $V\left(\hat{W}_2\right)$ ด้วย $v\left(\hat{W}_2\right)$ ดังนั้นค่าประมาณของ $V\left(\hat{\mu}_{pst}\right)$ มีค่าเท่ากับ

$$v\left[\hat{\mu}_{pst}\right] \simeq \left(1-\frac{n_T}{\hat{N}_T}\right)\frac{1}{n_T}\sum_{h=1}^{2}\hat{W}_h s_h^2 + \frac{1}{n_T^2}\sum_{h=1}^{2}\left(1-\hat{W}_h\right)s_h^2 +$$

$$\frac{1}{\hat{N}_T^2}\left[\hat{W}_1^2\hat{\mu}_1^2 + \left(\hat{W}_2 - 1\right)^2\hat{\mu}_2^2\right]v\left(\hat{N}_2\right)$$

เมื่อ

$$v\left(\hat{N}_2\right) = N_1^2\left(1-\frac{n_1}{N_1}\right)\frac{s_m^2}{n_1} \text{ (จากทฤษฎี 4)}$$

จากทฤษฎี 6 เป็นการประมาณค่าเฉลี่ยอย่างง่ายด้วยเทคนิคการแบ่งชั้นภูมิภายหลังการแจงนับ และใช้แผนการสุ่มตัวอย่างแบบง่ายไม่แทนที่ เมื่อเกิดปัญหากรอบตัวอย่างไม่ครอบคลุม ด้วยวิธี Half-open Interval แต่หากในกรอบตัวอย่างนั้นมีสาระสนเทศแหรือตัวแปรช่วยที่มีความสัมพันธ์กับตัวแปรที่ศึกษา และคาดหวังว่าการใช้ตัวแปรช่วยจะช่วยให้การประมาณ

ค่าเฉลี่ยมีความแม่นยำมากขึ้น เราจะใช้ตัวประมาณค่าที่ใช้ตัวแปรช่วย ได้แก่ ตัวประมาณอัตราส่วน (ดังทฤษฎี 2) และตัวประมาณการถดถอย (ดังทฤษฎี 3) จะได้ตัวประมาณอัตราส่วนและตัวประมาณการถดถอย ด้วยเทคนิคการแบ่งชั้นภูมิภายหลังการแจงนับ ดังนี้

ทฤษฎี 7 การประมาณค่าเฉลี่ยประชากรด้วยตัวประมาณอัตราส่วน เมื่อเกิดปัญหากรอบตัวอย่างไม่ครอบคลุม จากทฤษฎี 2 ประยุกต์ใช้กับทฤษฎี 6 จะได้ตัวประมาณอัตราส่วนภายหลังการแจงนับ ดังนี้

1. $\hat{\mu}_{Rpst}$ เป็นตัวประมาณค่าเฉลี่ยที่เอนเอียงของค่าเฉลี่ยประชากร μ_T

$$\hat{\mu}_{Rpst} = \hat{W}_1\hat{\mu}_{R1} + \hat{W}_2\hat{\mu}^*_{R2} \tag{35}$$

เมื่อ

$$\hat{R}_h = \frac{\hat{\mu}_h}{\hat{\mu}_{xh}},\ \hat{\mu}_{xh} = \frac{1}{n_h}\sum_{i=1}^{n_h} x_{hi} \text{ และ } \mu_{x1} = \frac{1}{N_1}\sum_{i=1}^{N_1} x_{1i}$$

โดยที่

$$\hat{\mu}_{R1} = \hat{R}_1\mu_{x1} \text{ และ } \hat{\mu}_{R2} = \hat{R}_2\mu_{x2} \approx \hat{\mu}^*_{R2} \text{ เมื่อ } \hat{\mu}^*_{R2} = \hat{R}_2\mu_{x1}$$

2 ความแปรปรวนแบบไม่มีเงื่อนไขของตัวประมาณค่าเฉลี่ย

$$V\left[\hat{\mu}_{Rpst}\right] \simeq \left(1-\frac{n_T}{\hat{N}_T}\right)\frac{1}{n_T}\sum_{h=1}^{2}\hat{W}_h S^2_{dh} + \frac{1}{n_T^2}\sum_{h=1}^{2}\left(1-\hat{W}_h\right)S^2_{dh}$$

$$+\frac{1}{N_T^2}\left[W_1^2\mu_1^2 + \left(W_2-1\right)^2\mu_2^2\right]V\left(\hat{N}_2\right) \tag{36}$$

เมื่อ

$$S^2_{dh} = \frac{1}{N_h-1}\sum_{i=1}^{N_h}(y_{hi} - R_h x_{hi})^2$$

โดยที่

$$R_h = \frac{\mu_h}{\mu_{hx}} \text{ และ } \mu_{hx} = \frac{1}{N_h}\sum_{i=1}^{N_h} x_{hi}$$

3. ค่าประมาณความแปรปรวนแบบไม่มีเงื่อนไขของตัวประมาณค่าเฉลี่ย

$$v\left[\hat{\mu}_{Rpst}\right] \simeq \left(1-\frac{n_T}{\hat{N}_T}\right)\frac{1}{n_T}\sum_{h=1}^{2}\hat{W}_h s_{dh}^2 + \frac{1}{n_T^2}\sum_{i=1}^{2}\left(1-\hat{W}_h\right)s_{dh}^2$$

$$+\frac{1}{\hat{N}_T^2}\left[\hat{W}_1^2\hat{\mu}_{R1}^2 + \left(\hat{W}_2 - 1\right)^2\hat{\mu}_{R2}^2\right]v\left(\hat{N}_2\right) \tag{37}$$

เมื่อ

$$s_{dh}^2 = \frac{1}{n_h - 1}\sum_{i=1}^{n_h}(y_{hi} - \hat{R}_h x_{hi})^2$$

ทฤษฎี 8 การประมาณค่าเฉลี่ยประชากรด้วยตัวประมาณการถดถอย เมื่อเกิดปัญหากรอบตัวอย่างไม่ครอบคลุม จากทฤษฎี 3 ประยุกต์ใช้กับทฤษฎี 6 จะได้ตัวประมาณการถดถอยภายหลังการแจงนับ ดังนี้

1. $\hat{\mu}_{Lpst}$ เป็นตัวประมาณค่าเฉลี่ยที่เอนเอียงของค่าเฉลี่ยประชากร μ_T

$$\hat{\mu}_{Lpst} = \hat{W}_1\hat{\mu}_{L1} + \hat{W}_2\hat{\mu}_{L2}^* \tag{38}$$

เมื่อ

$$\hat{\mu}_{L1} = \hat{\mu}_1 + \hat{\beta}_1(\mu_{1x} - \hat{\mu}_{x1}) \text{ และ } \hat{\mu}_{L2} = \hat{\mu}_2 + \hat{\beta}_2(\mu_{x2} - \hat{\mu}_{x2}) \approx \hat{\mu}_{L2}^*$$

$$\hat{\mu}^{*}_{L2} = \hat{\mu}_2 + \hat{\beta}_2 (\mu_{x1} - \hat{\mu}_{x2}) \text{ และ } \hat{\beta}_h = \frac{\sum_{i=1}^{n_h} (x_{hi} - \hat{\mu}_{xh})(y_{hi} - \hat{\mu}_h)}{\sum_{i=1}^{n_h} (x_{hi} - \hat{\mu}_{xh})^2}$$

2. ความแปรปรวนแบบไม่มีเงื่อนไขของตัวประมาณค่าเฉลี่ย

$$V\left[\hat{\mu}_{Lpst}\right] \simeq \left(1 - \frac{n_T}{\hat{N}_T}\right)\frac{1}{n_T}\sum_{h=1}^{2}\hat{W}_h S^2_{zh} + \frac{1}{n_T^2}\sum_{h=1}^{2}\left(1-\hat{W}_h\right)S^2_{zh}$$

$$+\frac{1}{N_T^2}\left[W_1^2\mu_1^2 + \left(W_2 - 1\right)^2 \mu_2^2\right]V\left(\hat{N}_2\right) \qquad (39)$$

เมื่อ

$$S^2_{zh} = \frac{1}{N_h - 1}\sum_{i=1}^{N_h}\left[\left(y_{hi} - \mu_h\right) - \beta_h\left(x_{hi} - \mu_{xh}\right)\right]^2$$

และ

$$\beta_h = \frac{\sum_{i=1}^{N_h}\left(x_{hi} - \mu_{hx}\right)\left(y_{hi} - \mu_h\right)}{\sum_{i=1}^{N_h}\left(x_{hi} - \mu_{hx}\right)^2}$$

3. ค่าประมาณความแปรปรวนแบบไม่มีเงื่อนไขของตัวประมาณค่าเฉลี่ย

$$v\left[\hat{\mu}_{Lpst}\right] \simeq \left(1 - \frac{n_T}{\hat{N}_T}\right)\frac{1}{n_T}\sum_{h=1}^{2}\hat{W}_h s^2_{zh} + \frac{1}{n_T^2}\sum_{h=1}^{2}\left(1-\hat{W}_h\right)s^2_{zh}$$

$$+\frac{1}{\hat{N}_T^2}\left[\hat{W}_1^2\hat{\mu}_{L1}^2 + \left(\hat{W}_2 - 1\right)^2 \hat{\mu}_{L2}^2\right]v\left(\hat{N}_2\right) \qquad (40)$$

เมื่อ

$$s_{zh}^2 = \frac{1}{n_h - 2}\left[\sum_{i=1}^{n_h}(y_{hi} - \hat{\mu}_h)^2 - \hat{\beta}_h^2 \sum_{i=1}^{n_h}(x_{hi} - \hat{\mu}_{xh})^2\right]$$

การสุ่มตัวอย่างแบบง่ายไม่แทนที่เมื่อเกิดปัญหากรอบตัวอย่างไม่ครอบคลุมด้วยวิธี **Half-open Interval** เพื่อให้เข้าใจง่ายจะประยุกต์ใช้กรอบตัวอย่างที่ไม่ครอบคลุมที่เสนอโดย Agarwal and Gupta (2008, p. 2) และสมมติค่าของตัวแปรที่ศึกษา y_{1i} และค่าของตัวแปรช่วย x_{1i} ซึ่ง i = 1, ..., 70 โดยกำหนดให้หมายเลข 1-70 คือหมายเลขแสดงลำดับของครัวเรือนที่อยู่ในกรอบตัวอย่าง และวงกลม ⊗ ระหว่างหมายเลขแสดงลำดับครัวเรือน คือ ครัวเรือนที่ไม่ได้อยู่ในกรอบตัวอย่าง และให้ y_{1i} เป็นค่าใช้จ่ายของครัวเรือน (หน่วย: พันบาท/เดือน) และ x_{1i} เป็นขนาดของครัวเรือน ตามลำดับ ตัวเลขที่อยู่ในวงเล็บคือ (x_{1i}, y_{1i}) ซึ่งจะได้ค่าของประชากร (แสดงดังตาราง 2)

1 (5, 5.6)	⊗ (3, 4.5)	⊗ (5, 6.3)	2 (5, 6.3)	3 (2, 3.3)	4 (3, 4.4)	⊗ (3, 4.6)	5 (5, 5.1)	6 (4, 5.4)	7 (4, 4.6)
8 (4 5.2)	9 (3, 4.0)	⊗ (5, 6.0)	10 (4, 4.8)	⊗ (7, 6.3)	11 (3, 4.8)	⊗ (2, 3.3)	⊗ (7, 7.4)	12 (4, 5.8)	13 (4, 4.6)
14 (3, 4.3)	15 (3, 4.3)	16 (4, 5.2)	⊗ (5, 6.1)	17 (3, 4.4)	18 (6, 6.3)	19 (4, 5.5)	⊗ (5, 5.8)	20 (4, 5.3)	21 (4, 4.9)
22 (4, 5.3)	23 (5, 6.1)	⊗ (4, 4.4)	⊗ (3, 4.8)	⊗ (3, 3.9)	24 (2, 3.2)	25 (5, 6.0)	26 (3, 4.7)	27 (4, 5.0)	28 (5, 5.7)
29 (4, 5.5)	30 (2, 3.7)	⊗ (3, 4.0)	31 (2, 3.8)	32 (5, 5.5)	33 (4, 4.3)	34 (3, 3.7)	35 (3, 4.7)	36 (4, 4.7)	37 (2, 4.1)
38 (2, 4.1)	39 (4, 5.2)	⊗ (3, 4.3)	⊗ (4, 5.5)	40 (2, 3.5)	⊗ (6, 6.5)	⊗ (5, 6.1)	41 (5, 6.1)	42 (5, 5.5)	⊗ (3, 4.6)
43 (4, 4.8)	44 (3, 5.0)	45 (3, 4.3)	46 (2, 4.2)	47 (4, 5.3)	48 (5, 5.7)	49 (5, 6.0)	50 (3, 4.9)	51 (5, 5.5)	52 (6, 6.6)
53 (6, 6.8)	54 (4, 4.7)	⊗ (2, 3.3)	⊗ (4, 5.2)	⊗ (3, 4.4)	55 (6, 6.8)	56 (5, 4.9)	⊗ (1, 3.6)	57 (4, 4.4)	⊗ (4, 5.4)
⊗ (4, 5.5)	⊗ (5, 5.7)	58 (4, 5.5)	59 (4, 5.0)	⊗ (4, 4.8)	⊗ (4, 4.7)	⊗ (4, 4.8)	60 (6, 5.9)	61 (5, 5.5)	62 (2, 3.8)
63 (1, 2.6)	64 (4, 5.4)	⊗ (5, 4.9)	65 (3, 4.2)	66 (5, 5.7)	67 (6, 6.3)	68 (4, 4.2)	⊗ (4, 5.2)	69 (5, 5.8)	70 (5, 5.1)

ภาพ 2 แสดงตัวอย่างการเกิดปัญหากรอบตัวอย่างไม่ครอบคลุม

ตาราง 2 แสดงค่าพารามิเตอร์ของประชากรสำหรับตัวอย่างกรอบตัวอย่างที่ไม่ครอบคลุม

รายละเอียด	ประชากรที่อยู่ในกรอบตัวอย่าง	ประชากรที่ไม่อยู่ในกรอบตัวอย่าง	ประชากร
จำนวน	$N_1 = 70$	$N_2 = 30$	$N_T = 100$
ค่าเฉลี่ยของตัวแปรที่ศึกษา	$\mu_1 = 4.9914$	$\mu_2 = 5.0366$	$\mu_T = 5.0130$
ค่าเฉลี่ยของตัวแปรช่วย	$\mu_{x1} = 3.9286$	$\mu_{x2} = 4.0333$	$\mu_{Tx} = 3.9600$

จากข้อมูลของประชากรจะได้ค่าสัมประสิทธิ์สหสัมพันธ์ของประชากร (ρ) เท่ากับ 0.902สำหรับค่าสัมประสิทธิ์สหสัมพันธ์ของประชากรที่อยู่ในกรอบตัวอย่าง (ρ_1) เท่ากับ 0.900 และค่าสัมประสิทธิ์สหสัมพันธ์ของประชากรที่ไม่อยู่ในกรอบตัวอย่าง (ρ_2) เท่ากับ 0.904 ดังนั้นตัวแปรที่ศึกษา และตัวแปรช่วยมีความสัมพันธ์กันสูง

ตัวอย่าง 1 สุ่มตัวอย่างแบบง่ายไม่แทนที่ 20% จากรายชื่อครัวเรือนในกรอบตัวอย่าง (ดังแสดงในภาพ 2) ขนาด $N_1 = 70$ ดังนั้นจะได้ตัวอย่างที่อยู่ในกรอบตัวอย่างขนาด $n_1 = 14$ ซึ่งยกตัวอย่างการสุ่มตัวอย่าง ดังนี้

สุ่มตัวอย่างหน่วยที่ 1 ได้ตัวอย่างลำดับที่ $i = 24$ (แรเงาดังแสดงในภาพ 2) และได้หน่วยตัวอย่างที่ไม่อยู่ในกรอบตัวอย่างระหว่างหน่วยตัวอย่างในกรอบตัวอย่างหน่วยที่ $i = 24$ และหน่วยที่ $i+1 = 25$ เท่ากับ 0 หน่วย ($m_{1(24)} = 0$) ต่อมาสุ่มตัวอย่างหน่วยที่ 2 ได้ตัวอย่างลำดับที่ $i = 64$ และได้หน่วยตัวอย่างที่ไม่อยู่ในกรอบตัวอย่างระหว่างหน่วยตัวอย่างในกรอบตัวอย่างหน่วยที่ $i = 64$ และหน่วยที่ $i+1 = 65$ เท่ากับ 1 หน่วย ($m_{1(64)} = 1$) ซึ่งจะสุ่มตัวอย่างในลักษณะนี้จนได้ตัวอย่างที่อยู่ในกรอบตัวอย่างครบ 14 หน่วย ดังนั้นจะได้ตัวอย่างที่อยู่ในกรอบตัวอย่าง $n_1 = 14$ และจะได้ขนาดตัวอย่างที่ไม่อยู่ในกรอบตัวอย่าง เท่ากับ $n_2 = \sum_{i=1}^{14} m_{1i} = 0 + 1 + \ldots + 0 + 0 = 7$ (ดังแสดงในตาราง 3) ดังนั้นจำนวนตัวอย่างทั้งหมดมีค่าเท่ากับ $n_T = n_1 + n_2 = 14 + 7 = 21$

ตาราง 3 แสดงตัวอย่างข้อมูลที่ได้จากการสุ่มตัวอย่าง เมื่อเกิดปัญหากรอบตัวอย่างไม่ครอบคลุม

ตัวอย่างที่อยู่ในกรอบตัวอย่าง (h = 1)				ตัวอย่างที่ไม่อยู่ในกรอบตัวอย่าง (h = 2)		
หน่วยที่ (i)	m_{1i}	y_{1i}	x_{1i}	**หน่วยที่**	y_{2i}	x_{2i}
24	0	3.2	2	64_1	4.9	5
64	1	5.4	4	68_1	5.2	4
62	0	3.8	2	11_1	3.3	2
6	0	5.4	4	11_2	7.4	7
43	0	4.8	4	54_1	3.3	2
68	1	4.2	4	54_2	5.2	4
27	0	5	4	54_3	4.4	3
11	2	4.8	3			
49	0	6	5			
54	3	4.7	4			
33	0	4.3	4			
41	0	6.1	5			
69	0	5.8	5			
15	0	4.3	3			
ผลรวม	7	67.8	54	**ผลรวม**	33.7	27

หมายเหตุ: หน่วยที่ 64_1 แทน ตัวอย่างที่ไม่อยู่ในกรอบตัวอย่างลำดับที่ 1 ที่อยู่ระหว่างหน่วยตัวอย่างที่อยู่ในกรอบตัวอย่างลำดับที่ i = 64 และ $i+1$ = 65

จากข้อมูลของตัวอย่าง จะได้ค่าสัมประสิทธิ์สหสัมพันธ์ของตัวอย่างทั้งหมด (r) เท่ากับ 0.912 สำหรับค่าสัมประสิทธิ์สหสัมพันธ์ของตัวอย่างที่อยู่ในกรอบตัวอย่าง (r_1) เท่ากับ 0.851 และค่าสัมประสิทธิ์สหสัมพันธ์ของตัวอย่างที่ไม่อยู่ในกรอบตัวอย่าง (r_2) เท่ากับ 0.962 ดังนั้นตัวแปรที่ศึกษาและตัวแปรช่วยมีความสัมพันธ์กันสูง

ตัวอย่าง 2 จากทฤษฎี 4 จะได้ตัวประมาณของจำนวนประชากรที่ไม่อยู่ในกรอบตัวอย่าง N_2 มีค่าเท่ากับ

$$\hat{N}_2 = N_1 \frac{1}{n_1} \sum_{i=1}^{n_1} m_{1i} = (70)\frac{1}{14}(0+1+...+0+0) = (70)\frac{1}{14}(7) = 35$$

และค่าประมาณความแปรปรวน มีค่าเท่ากับ

$$v\left(\hat{N}_2\right) = N_1^2\left(1-\frac{n_1}{N_1}\right)\frac{s_m^2}{n_1} = (70)^2\left(1-\frac{14}{70}\right)\frac{1.1731}{14} = 328.4680$$

เมื่อ

$$s_m^2 = \frac{1}{n_1-1}\sum_{i=1}^{n_1}(m_{1i}-\bar{m})^2 = \frac{1}{14-1}(15.25) = 1.1731$$

$$\bar{m} = \frac{1}{n_1}\sum_{i=1}^{n_1}m_{1i} = \frac{1}{14}(0+1+...+0+0) = \frac{1}{14}(7) = 0.50$$

ตัวอย่าง 3 จากทฤษฎี 5 จะได้ตัวประมาณที่ไม่เอนเอียงของจำนวนประชากร N_T มีค่าเท่ากับ

$$\hat{N}_T = N_1 + \hat{N}_2 = 70 + 35 = 105$$

และค่าประมาณความแปรปรวน มีค่าเท่ากับ

$$v\left(\hat{N}_T\right) = N_1^2\left(1-\frac{n_1}{N_1}\right)\frac{s_m^2}{n_1} = v\left(\hat{N}_2\right) = 328.4680$$

ตัวอย่าง 4 จากทฤษฎี 6 จะได้ตัวประมาณอย่างง่าย ของค่าเฉลี่ยประชากร μ_T มีค่าเท่ากับ

$$\hat{\mu}_{pst} = \hat{W}_1\hat{\mu}_1 + \hat{W}_2\hat{\mu}_2 = \frac{70}{105}\left(\frac{67.8}{14}\right) + \frac{35}{105}\left(\frac{33.7}{7}\right) = 4.8333$$

โดยคำนวณค่าต่าง ๆ ได้ดังนี้

$$\hat{\mu}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} y_{1i} = \frac{1}{14}(3.2+5.4+...+5.8+4.3) = \frac{1}{14}(67.8) = 4.8429$$

$$\hat{\mu}_2 = \frac{1}{n_2}\sum_{i=1}^{n_2} y_{2i} = \frac{1}{7}(4.9+5.2+...+5.2+4.4) = \frac{1}{7}(33.7) = 4.8143$$

$$\hat{W}_1 = \frac{N_1}{\hat{N}_T} = \frac{70}{105} \text{ และ } \hat{W}_2 = \frac{\hat{N}_2}{\hat{N}_T} = \frac{35}{105}$$

$$s_1^2 = \frac{1}{n_1 - 1}\sum_{i=1}^{n_1}(y_{1i} - \hat{\mu}_1)^2 = \frac{1}{14\text{-}1}(9.2943) = 0.7149$$

$$s_2^2 = \frac{1}{n_2 - 1}\sum_{i=1}^{n_2}(y_{2i} - \hat{\mu}_2)^2 = \frac{1}{7\text{-}1}(11.7486) = 1.9581$$

มีค่าเอนเอียง เท่ากับ

$$Bias(\hat{\mu}_{pst}) = \hat{\mu}_{pst} - \hat{\mu}_T = 4.8333 - 5.0130 = -0.1797$$

ค่าประมาณความแปรปรวนแบบไม่มีเงื่อนไขของตัวประมาณค่าเฉลี่ย มีค่าเท่ากับ

$$v\left[\hat{\mu}_{pst}\right] \simeq \left(1 - \frac{n_T}{\hat{N}_T}\right)\frac{1}{n_T}\sum_{h=1}^{2}\hat{W}_h s_h^2 + \frac{1}{n_T^2}\sum_{h=1}^{2}\left(1 - \hat{W}_h\right)s_h^2$$

$$+ \frac{1}{\hat{N}_T^2}\left[\hat{W}_1^2\hat{\mu}_1^2 + \left(\hat{W}_2 - 1\right)^2\hat{\mu}_2^2\right] v\left(\hat{N}_2\right)$$

$$\simeq \left(1\text{-}\frac{21}{105}\right)\frac{1}{21}\left[\frac{70}{105}(0.7149) + \frac{35}{105}(1.9518)\right] +$$

$$\frac{1}{(21)^2}\left[\left(1\text{-}\frac{70}{105}\right)0.7149 + \left(1\text{-}\frac{35}{105}\right)1.9581\right] +$$

$$\frac{1}{(105)^2}\left[\left(\frac{70}{105}\right)^2(4.8429)^2+\left(\frac{35}{105}\right)^2(4.8143)^2\right](328.4680)$$

$$\simeq 0.0429 + 0.0035 + 0.3873 = 0.4337$$

และค่าความคลาดเคลื่อนกำลังสองเฉลี่ย เท่ากับ

$$MSE(\hat{\mu}_{pst}) = \left[Bias(\hat{\mu}_{pst})\right]^2 + v\left(\hat{\mu}_{pst}\right) = (-0.1797)^2 + 0.4337 = 0.4660$$

ตัวอย่าง 5 จากทฤษฎี 7 จะได้ตัวประมาณอัตราส่วน ของค่าเฉลี่ยประชากร μ_T มีค่าเท่ากับ

$$\hat{\mu}_{Rpst} = \hat{W}_1\hat{\mu}_{R1} + \hat{W}_2\hat{\mu}^*_{R2} = \frac{70}{105}(4.9326) + \frac{35}{105}(4.9035) = 4.9229$$

โดยคำนวณค่าต่าง ๆ ได้ดังนี้

$$\hat{\mu}_{x1} = \frac{1}{n_1}\sum_{i=1}^{n_1} x_{1i} = \frac{1}{14}(1+3+...+7+3) = \frac{1}{14}(54) = 3.8571$$

$$\hat{\mu}_{x2} = \frac{1}{n_1}\sum_{i=1}^{n_1} x_{1i} = \frac{1}{7}(5+4+...+4+3) = \frac{1}{7}(27) = 3.8571$$

$$\hat{R}_1 = \frac{\hat{\mu}_1}{\hat{\mu}_{x1}} = \frac{67.8/14}{54/14} = \frac{67.8}{54} = 1.2556$$

$$\hat{R}_2 = \frac{\hat{\mu}_2}{\hat{\mu}_{x2}} = \frac{33.7/7}{27/7} = \frac{33.7}{27} = 1.2481$$

$$\mu_{x1} = \frac{1}{N_1}\sum_{i=1}^{N_1} x_{1i} = \frac{1}{70}(275) = 3.9286$$

$$\hat{\mu}_{R1} = \hat{R}_1\mu_{x1} = \frac{67.8}{54}(3.9286) = 4.9326$$

$$\hat{\mu}^{*}_{R2} = \hat{R}_2\mu_{x1} = \frac{33.7}{27}(3.9286) = 4.9035$$

$$\hat{\mu}_{R2} = \hat{R}_2\mu_{x2} \approx \hat{\mu}^{*}_{R2} = 4.9035$$

$$s^2_{d1} = \frac{1}{n_1 - 1}\sum_{i=1}^{n_1}(y_{1i} - \hat{R}_1 x_{1i})^2 = \frac{1}{14\text{-}1}(8.2380) = 0.6337$$

$$s^2_{d2} = \frac{1}{n_2 - 1}\sum_{i=1}^{n_2}(y_{2i} - \hat{R}_2 x_{2i})^2 = \frac{1}{7\text{-}1}(5.3929) = 0.8988$$

มีค่าเอนเอียง เท่ากับ

$$Bias(\hat{\mu}_{Rpst}) = \hat{\mu}_{Rpst} - \hat{\mu}_T = 4.9229 - 5.0130 = 0.0901$$

ค่าประมาณความแปรปรวนแบบไม่มีเงื่อนไขของตัวประมาณค่าเฉลี่ย มีค่าเท่ากับ

$$v\left[\hat{\mu}_{Rpst}\right] \simeq \left(1 - \frac{n_T}{\hat{N}_T}\right)\frac{1}{n_T}\sum_{h=1}^{2}\hat{W}_h s^2_{dh} + \frac{1}{n_T^2}\sum_{h=1}^{2}\left(1 - \hat{W}_h\right)s^2_{dh}$$

$$+ \frac{1}{\hat{N}_T^2}\left[\hat{W}_1^2\hat{\mu}^2_{R1} + \left(\hat{W}_2 - 1\right)^2\hat{\mu}^2_{R2}\right]v\left(\hat{N}_2\right)$$

$$\simeq \left(1\text{-}\frac{21}{105}\right)\frac{1}{21}\left[\frac{70}{105}(0.6337) + \frac{35}{105}(0.8988)\right] +$$

$$\frac{1}{(21)^2}\left[\left(1\text{-}\frac{70}{105}\right)0.6337 + \left(1\text{-}\frac{35}{105}\right)0.8988\right] +$$

$$\frac{1}{(105)^2}\left[\left(\frac{70}{105}\right)^2(4.9326)^2+\left(\frac{35}{105}\right)^2(4.9035)^2\right](328.4680)$$

$$\simeq 0.0275+0.0018+0.4018 = 0.4331$$

และค่าความคลาดเคลื่อนกำลังสองเฉลี่ย เท่ากับ

$$MSE(\hat{\mu}_{Rpst}) = \left[Bias(\hat{\mu}_{Rpst})\right]^2 + v\left(\hat{\mu}_{Rpst}\right) = (-0.0901)^2+0.4331 = 0.4412$$

ตัวอย่าง 6 จากทฤษฎี 8 จะได้ตัวประมาณการถดถอยของค่าเฉลี่ยประชากร μ_T มีค่าเท่ากับ

$$\hat{\mu}_{Lpst} = \hat{W}_1\hat{\mu}_{L1} + \hat{W}_2\hat{\mu}^*_{L2}\left(\frac{70}{105}\right)4.8896+\left(\frac{35}{105}\right)4.8685 = 4.8826$$

โดยคำนวณค่าต่าง ๆ ได้ดังนี้

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n_1}(x_{1i}-\hat{\mu}_{x1})(y_{1i}-\hat{\mu}_1)}{\sum_{i=1}^{n_1}(x_{1i}-\hat{\mu}_{x1})^2} = \frac{10.2857}{15.7143} = 0.6545$$

$$\hat{\beta}_2 = \frac{\sum_{i=1}^{n_2}(x_{2i}-\hat{\mu}_{x2})(y_{2i}-\hat{\mu}_2)}{\sum_{i=1}^{n_2}(x_{2i}-\hat{\mu}_{x2})^2} = \frac{14.3143}{18.8571} = 0.7591$$

$$\hat{\mu}_{L1} = \hat{\mu}_1 + \hat{\beta}_1\left(\mu_{1x}-\hat{\mu}_{x1}\right) = \frac{67.8}{14}+0.6545\left(3.9286-\frac{54}{14}\right) = 4.8896$$

$$\hat{\mu}^*_{L2} = \hat{\mu}_2 + \hat{\beta}_2\left(\mu_{x1}-\hat{\mu}_{x2}\right) = \frac{33.7}{7}+0.7591\left(3.9286-\frac{27}{7}\right) = 4.8685$$

$$\hat{\mu}_{L2} = \hat{\mu}_2 + \hat{\beta}_2\left(\mu_{x2} - \hat{\mu}_{x2}\right) \approx \hat{\mu}^{*}_{L2} = 4.8685$$

$$s^2_{z1} = \frac{1}{n_1 - 2}\left[\sum_{i=1}^{n_1}(y_{1i} - \hat{\mu}_1)^2 + \hat{\beta}_1^2\sum_{i=1}^{n_1}(x_{1i} - \hat{\mu}_{x1})^2\right]$$

$$= \frac{1}{14\text{-}2}[9.2943\text{-}(0.6545)^2(15.7143)] = 0.2136$$

$$s^2_{z2} = \frac{1}{n_2 - 2}\left[\sum_{i=1}^{n_2}(y_{2i} - \hat{\mu}_2)^2 + \hat{\beta}_2^2\sum_{i=1}^{n_2}(x_{2i} - \hat{\mu}_{x2})^2\right]$$

$$= \frac{1}{7\text{-}2}[11.7486\text{-}(0.7591)^2(18.8571)] = 0.1765$$

มีค่าเอนเอียง เท่ากับ

$$Bias(\hat{\mu}_{Lpst}) = \hat{\mu}_{Lpst} - \hat{\mu}_T = 4.8826 - 5.0130 = -0.1304$$

ค่าประมาณความแปรปรวนแบบไม่มีเงื่อนไขของตัวประมาณค่าเฉลี่ย มีค่าเท่ากับ

$$v\left[\hat{\mu}_{Lpst}\right] \simeq \left(1 - \frac{n_T}{\hat{N}_T}\right)\frac{1}{n_T}\sum_{h=1}^{2}\hat{W}_h s^2_{zh} + \frac{1}{n_T^2}\sum_{h=1}^{2}\left(1 - \hat{W}_h\right)s^2_{zh}$$

$$+\frac{1}{\hat{N}_T^2}\left[\hat{W}_1^2\hat{\mu}_{L1}^2 + \left(\hat{W}_2 - 1\right)^2\hat{\mu}_{L2}^2\right]v\left(\hat{N}_2\right)$$

$$\simeq \left(1\text{-}\frac{21}{105}\right)\frac{1}{21}\left[\frac{70}{105}(0.2136) + \frac{35}{105}(0.1765)\right] +$$

$$\frac{1}{(21)^2}\left[\left(1\text{-}\frac{70}{105}\right)0.2136 + \left(1\text{-}\frac{35}{105}\right)0.1765\right] +$$

$$\frac{1}{(105)^2}\left[\left(\frac{70}{105}\right)^2(4.8896)^2+\left(\frac{35}{105}\right)^2(4.8685)^2\right](328.4680)$$

$$\simeq 0.0077 + 0.0004 + 0.3950 = 0.4031$$

และค่าความคลาดเคลื่อนกำลังสองเฉลี่ย เท่ากับ

$$MSE(\hat{\mu}_{Lpst}) = \left[Bias(\hat{\mu}_{Lpst})\right]^2 + v\left(\hat{\mu}_{Lpst}\right) = (-0.1304)^2 + 0.4212 = 0.4382$$

จากตัวอย่าง 4-6 จะได้

$$MSE(\hat{\mu}_{Lpst}) = 0.4382 < MSE(\hat{\mu}_{Rpst}) = 0.4412 < MSE(\hat{\mu}_{pst}) = 0.4660$$

การไม่ได้รับความร่วมมือ

ความคลาดเคลื่อนที่ไม่ได้เกิดจากการสุ่มอีกประเภทหนึ่งที่ควรระมัดระวัง คือ ความคลาดเคลื่อนจากการไม่ได้รับความร่วมมือ ซึ่งเกิดขึ้นจากการไม่สามารถรวบรวมค่าของตัวแปรจากบางหน่วยตัวอย่างด้วยเหตุผลต่าง ๆ เช่น การปฏิเสธที่จะให้ความร่วมมือ การไม่สามารถหาที่อยู่ผู้ให้ข้อมูล ผู้ให้ข้อมูลไม่อยู่บ้านหรือผู้ให้ข้อมูลไม่สามารถตอบคำถามบางข้อได้ แบบสอบถามสูญหายภายหลังการสัมภาษณ์ การจงใจของพนักงานที่สอบถามไม่ครบทุกหน่วยย่อย เป็นต้น สำหรับความคลาดเคลื่อนจากการไม่ได้รับความร่วมมือนั้น อาจเป็นการไม่ได้รับความร่วมมือในการตอบข้อถามใดเลยของผู้ตอบ หรือหน่วยตัวอย่าง เรียกว่า การไม่ได้รับความร่วมมือในการตอบข้อถามบางหน่วย (Unit Nonresponse) หรือการไม่ได้รับความร่วมมือในการตอบข้อถามบางข้อเรียกว่า การไม่ได้รับความร่วมมือในการตอบข้อถามบางข้อ (Item Nonresponse) ในการวิจัยครั้งนี้ผู้วิจัยศึกษาค่าเฉลี่ยของตัวแปรที่ศึกษา Y และในกรอบตัวอย่างมีตัวแปรช่วย X ซึ่งทราบค่า ดังนั้นจะเก็บรวบรวมตัวแปรที่ศึกษา Y จากตัวอย่าง ดังนั้นหากไม่ได้รับความร่วมมือหรือไม่ได้ค่าของตัวแปรที่ศึกษา Y จะเป็นการไม่ได้รับความร่วมมือในการการตอบข้อถามบางหน่วย

1. **ความคลาดเคลื่อนจากการไม่ได้รับความร่วมมือ**

ในการสำรวจ หากไม่ได้รับความร่วมมือในการตอบข้อถามจำนวนมาก จะมีผลต่อค่าประมาณจากการสำรวจคือ ความเอนเอียงของค่าประมาณ ซึ่งมีค่าเพิ่มขึ้นเมื่ออัตราการไม่ให้ความร่วมมือสูงขึ้น (ประชุม สุวัตถี, 2552, หน้า 500) ทำให้การประมาณค่ามีความเอนเอียงเช่นเดียวกับความคลาดเคลื่อนจากกรอบตัวอย่าง ซึ่งปัญหาการไม่ได้รับความร่วมมือมักหลีกเลี่ยงไม่ได้ ถึงแม้ว่าผู้วิจัยจะสรรหาวิธีการแก้ไขปัญหาดังกล่าวแล้ว แต่ท้ายที่สุดก็ยังเกิดปัญหาขึ้นได้ ดังนั้นผู้วิจัยจำเป็นต้องอาศัยเทคนิคอื่น ๆ ที่ใช้ในการแก้ไข หรือลดปัญหาการไม่ได้รับความร่วมมือของหน่วยตัวอย่าง (ทวีศักดิ์ ศิริพรไพบูลย์, 2549, หน้า 202-204)

2. **แนวทางแก้ไขปัญหาการไม่ได้รับความร่วมมือ**

แนวทางแก้ไขปัญหาการไม่ได้รับความร่วมมือ (ประชุม สุวัตถี, 2552, หน้า 501) อาจจำแนกได้ดังนี้

2.1 ก่อนและระหว่างการเก็บรวบรวมข้อมูล ควรกำหนดสัดส่วนการไม่ได้รับความร่วมมือให้อยู่ในระดับต่ำที่จะยอมรับได้ เพื่อให้สัดส่วนการไม่ได้รับความร่วมมือที่ยังเหลืออยู่ ไม่มีผลหรือมีผลน้อยต่อการนำข้อมูลไปใช้ในการประมาณค่าพารามิเตอร์ ซึ่งมีเทคนิค 3 ประการที่ควรนำมาใช้ ได้แก่ การวางแผนการสำรวจให้เหมาะสม การกลับไปเก็บรวบรวมข้อมูลที่เก็บในรอบแรกไม่ได้ (Callbacks) และการติดตามสัมภาษณ์เพิ่มเติมทางจดหมาย และทางโทรศัพท์

2.2 ใช้เทคนิคพิเศษในการเก็บรวบรวมข้อมูลและการประมาณค่า เพื่อให้สามารถประมาณพารามิเตอร์ได้โดยไม่เอนเอียง เทคนิคที่ใช้ ได้แก่ การสุ่มตัวอย่างย่อยในกลุ่มตัวอย่างที่ไม่ให้ความร่วมมือ (Nonresponsdent Subsampling) และการเลือกแทนด้วยส่วนที่เก็บได้แล้วอย่างสุ่ม (Randomized Response)

2.3 ใช้ความสัมพันธ์ระหว่างตัวแปร และคำตอบที่เก็บได้ เพื่อสร้างตัวประมาณของส่วนที่ไม่ได้รับความร่วมมือ ได้แก่ การปรับค่าถ่วงน้ำหนักให้เหมาะสม หรือการประมาณค่าที่เก็บไม่ได้หรือการแทนที่ข้อมูลที่สูญหายด้วยวิธีที่เหมาะสม

3. **การสุ่มตัวอย่างย่อยในกลุ่มที่ไม่ให้ความร่วมมือ**

การใช้ตัวแปรช่วยร่วมในการประมาณค่าไม่เหมาะสมสำหรับกรณีที่หน่วยที่ไม่ให้ความร่วมมือที่ละทิ้งไม่ได้ ส่วนวิธีการปรับค่าถ่วงน้ำหนัก และการแทนค่าข้อมูลสูญหาย ทั้งสองวิธีมีเป้าหมายในการลดความเอนเอียง (Biased) จากปัญหาการไม่ได้รับความร่วมมือ แต่อย่างไรก็ตามวิธีการปรับค่าถ่วงน้ำหนัก และการแทนที่ข้อมูลสูญหายจะขึ้นอยู่กับข้อตกลงเกี่ยวกับกลไกการให้ความร่วมมือ ซึ่งไม่สามารถควบคุมได้เมื่อกลไกการให้ความร่วมมือที่สมมตินั้นไม่ถูกต้อง แล้วผลลัพธ์ที่ได้จากการประมาณค่าจะเอนเอียงอย่างรุนแรง ดังนั้นการประมาณค่าที่

เหมาะสมเมื่อเกิดการสูญหายแบบไม่สามารถละทิ้งได้ อาจต้องใช้สาระโดยตรงของกลุ่มที่ไม่ให้ความร่วมมือ หรือสาระโดยอ้อมที่เหมาะสมของกลุ่มที่ให้ความร่วมมือในการประมาณค่าสูญหาย ซึ่งแหล่งข้อมูลโดยส่วนใหญ่อันเป็นสาระโดยตรงจากกลุ่มที่ไม่ให้ความร่วมมือ ก็คือข้อมูลที่ได้จากตัวอย่างของกลุ่มที่ไม่ให้ความร่วมมือนั่นเอง และวิธีการที่จะได้มาของข้อมูลเหล่านั้นก็จะเป็นวิธีการที่ใช้ความพยายามในการเก็บรวบรวมข้อมูลของกลุ่มบุคคลดังกล่าวที่ไม่ให้ความร่วมมือให้กลายเป็นกลุ่มที่ให้ความร่วมมือ เช่น ใช้การเก็บข้อมูลทางไปรษณีย์ การสัมภาษณ์ทางโทรศัพท์ การสัมภาษณ์ด้วยตนเองในสถานที่ที่ผู้ให้ข้อมูลสะดวกในการให้ความร่วมมือ ซึ่งวิธีการดังกล่าวเสนอโดย Hansen and Hurwitz (1946) ถึงแม้ว่าการสุ่มตัวอย่างย่อยในกลุ่มที่ไม่ให้ความร่วมมืออาจต้องใช้ค่าใช้จ่าย และระยะเวลาในการดำเนินงานเพิ่มขึ้น กระนั้นถ้าปัญหาความเอนเอียงที่เกิดจากการไม่ได้รับความร่วมมือนี้รุนแรง วิธีการสุ่มตัวอย่างย่อยในกลุ่มที่ไม่ให้ความร่วมมือก็เป็นทางเลือกหนึ่งที่ดำเนินการแก้ไขปัญหาโดยไม่ต้องทำการสำรวจใหม่ให้ได้รับความร่วมมือทั้งหมด อันจะทำให้ต้องเสียเวลาและค่าใช้จ่ายมากกว่า และผลที่ได้จากการสุ่มตัวอย่างย่อยในกลุ่มที่ไม่ให้ความร่วมมือนี้ก็นำมาซึ่งตัวแทนของกลุ่มที่ไม่ให้ความร่วมมือ เพื่อนำไปประมาณค่าตัวแปรที่ศึกษาในกลุ่มที่ไม่ให้ความร่วมมือ เช่น Siripornpibul (2001) ได้เสนอการประมาณค่าเฉลี่ยด้วยตัวประมาณอย่างง่าย สำหรับการสุ่มตัวอย่างย่อยกลุ่มที่ไม่ให้ความร่วมมือ 2 ครั้ง และ Okafor and Lee (2000) เสนอการประมาณค่าเฉลี่ยด้วยตัวประมาณอัตราส่วน และตัวประมาณการถดถอยในแผนการสุ่มตัวอย่างแบบสองขั้นตอน เมื่อมีการสุ่มตัวอย่างย่อยในกลุ่มที่ไม่ให้ความร่วมมือ โดยมีเงื่อนไขว่าไม่ทราบตัวแปรช่วยที่คาดว่ามีความสัมพันธ์กับตัวแปรที่ศึกษาจึงสุ่มตัวอย่าง 2 ครั้ง โดยครั้งแรกทำการสุ่มตัวอย่างเพื่อศึกษาตัวแปรช่วย จากนั้นจึงสุ่มตัวอย่างครั้งที่สองซึ่งเป็นการสุ่มซ้ำบางส่วนของตัวอย่างในครั้งแรกเพื่อศึกษาตัวแปรที่ศึกษา และในการสุ่มตัวอย่างครั้งที่สองนี้เกิดปัญหาการไม่ได้รับความร่วมมือจากบางหน่วยของตัวอย่าง จึงทำการสุ่มตัวอย่างย่อยกลุ่มตัวอย่างที่ไม่ให้ความร่วมมืออีกครั้ง

4. การประมาณค่าเฉลี่ยประชากร เมื่อเกิดปัญหาการไม่ได้รับความร่วมมือด้วยการสุ่มตัวอย่างย่อยในกลุ่มที่ไม่ให้ความร่วมมือ

การประมาณค่าเฉลี่ยประชากร μ จากการสุ่มตัวอย่างแบบง่ายไม่แทนที่ขนาด n จากประชากรในกรอบตัวอย่าง N หน่วย กำหนดให้ n_r เป็นขนาดตัวอย่างที่ให้ความร่วมมือ และ n_{nr} เป็นขนาดตัวอย่างที่ไม่ให้ความร่วมมือ จากนั้นจึงสุ่มตัวอย่างย่อยในกลุ่มที่ไม่ความร่วมมือขนาด $n'_{nr} = n_{nr}/k$ เมื่อ k เป็นค่าคงที่ที่กำหนดไว้ล่วงหน้า $(k > 1)$ ก่อนที่จะนำเสนอทฤษฎีเกี่ยวกับ การสุ่มตัวอย่างย่อยในกลุ่มที่ไม่ให้ความร่วมมือ จะพิจารณาจาก 2 บทแทรก (Lemma) ของ Cochran (1977 as cited in Siripornpibul, 2001, p. 65) ดังนี้

บทแทรก 1 ความแปรปรวนตัวอย่าง (Sample Variance)

1. s_{rr}^2 เป็นค่าประมาณความแปรปรวนประชากร S^2
2. s_{nr}^2 เป็นค่าประมาณความแปรปรวนประชากรที่ไม่ให้ความร่วมมือ S_{nr}^2

บทแทรก 2 ค่าคาดหวัง (Expectation) ในการสุ่มตัวอย่าง 2 ครั้ง (Two-phase Sampling) จะได้ค่าคาดหวังของตัวประมาณ $\hat{\theta}$ ในการสุ่มตัวอย่าง 2 ครั้ง มีค่าเท่ากับ

$$E\left(\hat{\theta}\right) = E_1 E_2\left(\hat{\theta}\right)$$

เมื่อ E_1 และ E_2 คือ ค่าคาดหวังของพารามิเตอร์ θ สำหรับตัวอย่าง ขนาด n และค่าคาดหวังของพารามิเตอร์ θ สำหรับการสุ่มตัวอย่างย่อยในกลุ่มที่ไม่ให้ความร่วมมือ ขนาด n'_{nr} ตามลำดับ และค่าความแปรปรวน (Variance) ของตัวประมาณ $\hat{\theta}$ ในการสุ่มตัวอย่าง 2 ครั้ง มีค่าเท่ากับ

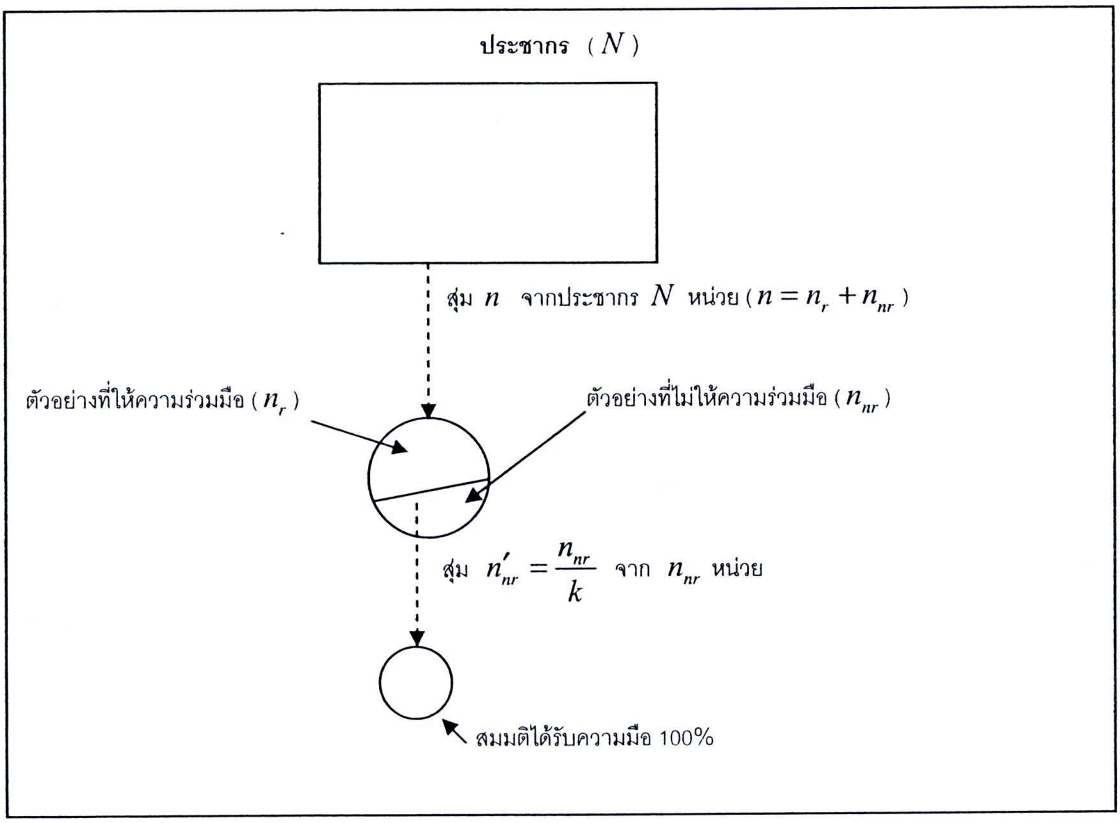
$$V\left(\hat{\theta}\right) = V_1 E_2\left(\hat{\theta}\right) + E_1 V_2\left(\hat{\theta}\right)$$

เมื่อ V_1 และ V_2 คือ ความแปรปรวนของพารามิเตอร์ θ สำหรับตัวอย่าง ขนาด n และค่าคาดหวังของพารามิเตอร์ θ สำหรับการสุ่มตัวอย่างย่อยในกลุ่มที่ไม่ให้ความร่วมมือ ขนาด n'_{nr} ตามลำดับ

การสุ่มตัวอย่างแบบง่ายไม่แทนที่ขนาด n จากประชากรในกรอบตัวอย่าง N หน่วย เมื่อเกิดปัญหาการไม่ได้รับความร่วมมือ และมีการสุ่มตัวอย่างย่อยในกลุ่มตัวอย่างที่ไม่ให้ความร่วมมือ (ดังแสดงในภาพ 3) มีขั้นตอนดังนี้

ขั้นตอน 1 สุ่มตัวอย่างแบบง่ายไม่แทนที่ขนาด n จากประชากรในกรอบตัวอย่าง N หน่วย กำหนดให้ n_r เป็นขนาดตัวอย่างที่ให้ความร่วมมือ และ n_{nr} เป็นขนาดตัวอย่างที่ไม่ให้ความร่วมมือ

ขั้นตอน 2 สุ่มตัวอย่างย่อยในกลุ่มที่ไม่ความร่วมมือขนาด $n'_{nr} = n_{nr}/k$ จากตัวอย่างที่ไม่ให้ความร่วมมือ n_{nr} หน่วย



ภาพ 3 แสดงการสุ่มตัวอย่างเมื่อเกิดปัญหาการไม่ได้รับความร่วมมือ

และกำหนดให้

N_{nr} แทน ขนาดประชากรที่ไม่ให้ความร่วมมือจากประชากรขนาด N

W_{nr} แทน สัดส่วนประชากรที่ไม่ให้ความร่วมมือ ; $W_{nr} = N_{nr} / N$

w_{nr} แทน สัดส่วนตัวอย่างที่ไม่ให้ความร่วมมือ ; $w_{nr} = n_{nr} / n$

n_{rr} แทน ขนาดตัวอย่างที่ให้ความร่วมมือทั้งหมด ; $n_{rr} = n_r + n'_{nr}$

y_r แทน ผลรวมของตัวแปรที่ศึกษาของตัวอย่างที่ให้ความร่วมมือ ;

$$y_r = \sum_{i=1}^{n_r} y_i$$

y_{nr} แทน ผลรวมของตัวแปรที่ศึกษาของตัวอย่างที่ไม่ให้ความร่วมมือ ;

$$y_{nr} = \sum_{i=1}^{n_{nr}} y_i$$

y แทน ผลรวมของตัวแปรที่ศึกษาของตัวอย่างขนาด n ;

$$y = \sum_{i=1}^{n} y_i = y_r + y_{nr}$$

y'_{nr} แทน ผลรวมของตัวแปรที่ศึกษาของตัวอย่างที่ให้ความร่วมมือในการสุ่มตัวอย่างย่อยในกลุ่มที่ไม่ให้ความร่วมมือ ; $y'_{nr} = \sum_{i=1}^{n'_{nr}} y_i$

x_r แทน ผลรวมของตัวแปรช่วยของตัวอย่างที่ให้ความร่วมมือ ; $x_r = \sum_{i=1}^{n_r} x_i$

x'_{nr} แทน ผลรวมของตัวแปรช่วยของตัวอย่างที่ให้ความร่วมมือจากการสุ่มตัวอย่างย่อยในกลุ่มที่ไม่ให้ความร่วมมือ ; $x'_{nr} = \sum_{i=1}^{n'_{nr}} x_i$

S^2_{nr} แทน ความแปรปรวนประชากรที่ไม่ให้ความร่วมมือ

s^2_{rr} แทน ความแปรปรวนตัวอย่างของกลุ่มตัวอย่างที่ให้ความร่วมมือ ขนาด n_{rr}

s^2_{nr} แทน ความแปรปรวนตัวอย่างของกลุ่มตัวอย่างที่ให้ความร่วมมือ ขนาด n'_{nr} จากการสุ่มตัวอย่างย่อยในกลุ่มที่ไม่ให้ความร่วมมือ ขนาด n_{nr}

การประมาณค่าเฉลี่ยประชากร μ จากการสุ่มตัวอย่างแบบง่ายไม่แทนที่ขนาด n จากประชากรในกรอบตัวอย่าง N หน่วย เมื่อเกิดปัญหาการไม่ได้รับความร่วมมือ และมีการสุ่มตัวอย่างย่อยในกลุ่มตัวอย่างที่ไม่ให้ความร่วมมือ มีรายละเอียดดังนี้

ทฤษฎี 9 การประมาณค่าเฉลี่ยประชากรด้วยตัวประมาณอย่างง่าย เมื่อเกิดปัญหาการไม่ได้รับความร่วมมือ

1. $\hat{\mu}^*$ เป็นตัวประมาณค่าเฉลี่ยที่ไม่เอนเอียงของค่าเฉลี่ยประชากร μ

$$\hat{\mu}^* = \frac{1}{n}\left(y_r + k\, y'_{nr}\right) \text{ เมื่อ } k = \frac{n_{nr}}{n'_{nr}} \tag{41}$$

2. ความแปรปรวนของตัวประมาณค่าเฉลี่ย

$$V\left(\hat{\mu}^*\right) = \left(1 - \frac{n}{N}\right)\frac{S^2}{n} + \frac{(k-1)}{n} W_{nr} S^2_{nr} \tag{42}$$

เมื่อ

$$S^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i-\mu)^2 \text{ โดยที่ } \mu = \frac{1}{N}\sum_{i=1}^{N}y_i$$

$$S_{nr}^2 = \frac{1}{N_{nr}-1}\sum_{i=1}^{N_{nr}}(y_i-\mu_{nr})^2 \text{ โดยที่ } \mu_{nr} = \frac{1}{N_{nr}}\sum_{i=1}^{N_{nr}}y_i$$

3. ค่าประมาณความแปรปรวนตัวประมาณค่าเฉลี่ย

$$v\left(\hat{\mu}^*\right) = \left(1-\frac{n}{N}\right)\frac{s_{rr}^2}{n} + \frac{(k-1)}{n}w_{nr}s_{nr}^2 \tag{43}$$

เมื่อ

$$s_{rr}^2 = \frac{1}{n_{rr}-1}\left[\sum_{i=1}^{n_{rr}}(y_i-\hat{\mu}_{rr})^2\right] \text{ โดยที่ } \hat{\mu}_{rr} = \frac{1}{n_{rr}}(y_r + y'_{nr})$$

$$s_{nr}^2 = \frac{1}{n'_{nr}-1}\sum_{i=1}^{n'_{nr}}(y_i-\hat{\mu}'_{nr})^2 \text{ โดยที่ } \hat{\mu}'_{nr} = \frac{y'_{nr}}{n'_{nr}}$$

พิสูจน์ แนวทางการพิสูจน์ดัง Siripornpibul (2001, pp. 66-68) เนื่องจากมีการสุ่มตัวอย่าง 2 ครั้ง ดังนั้นจะได้

$$E\left(\hat{\mu}^*\right) = E_1E_2\left(\hat{\mu}^*\right)$$

และ

$$V\left(\hat{\mu}^*\right) = V_1E_2\left(\hat{\mu}^*\right) + E_1V_2\left(\hat{\mu}^*\right)$$

1. เนื่องจาก

$$E\left(\hat{\mu}^*\right) = E_1E_2\frac{1}{n}\left[y_r + \frac{n_{nr}}{n'_{nr}}(y'_{nr})\right]$$

$$= E_1 \frac{1}{n}\left[y_r + n_{nr} E_2\left(\hat{\mu}'_{nr}\right)\right]$$

$$= E_1 \frac{1}{n}\left[y_r + n_{nr}\hat{\mu}_{nr}\right]$$

$$= E_1 \frac{1}{n}\left[y_r + y_{nr}\right]$$

$$= E_1\left[\hat{\mu}\right]$$

$$= \mu \tag{44}$$

ดังนั้น $\hat{\mu}^*$ เป็นตัวประมาณที่ไม่เอนเอียงของ μ

2. เนื่องจาก

$$V_1 E_2\left(\hat{\mu}^*\right) = V\left[\hat{\mu}\ \right] = \left(1-\frac{n}{N}\right)\frac{S^2}{n} \tag{45}$$

และ

$$E_1 V_2\left(\hat{\mu}^*\right) \ = E_1\left[(\frac{n_{nr}}{n})^2 V_2\left(\hat{\mu}'_{nr}\right)\right]$$

$$= E_1\left[\frac{n_{nr}^2}{n^2}\left(1-\frac{n'_{nr}}{n_{nr}}\right)\frac{S_{nr}^2}{n'_{nr}}\right]$$

$$= E_1\left[\frac{1}{n}\frac{n_{nr}}{n}\left(\frac{n_{nr}}{n'_{nr}}-1\right)S_{nr}^2\right]$$

$$= E_1\left[\frac{1}{n} w_{nr}\left(k-1\right)S_{nr}^2\right] \ ; \ k = \frac{n_{nr}}{n'_{nr}}$$

เมื่อ

$$E(w_{nr}) = E\left(\frac{n_{nr}}{n}\right) = \frac{N_{nr}}{N} = W_{nr}$$

จะได้

$$E_1V_2\left(\hat{\mu}^*\right) = \frac{(k-1)}{n}W_{nr}S_{nr}^2 \qquad (46)$$

ดังนั้นจะได้

$$V\left(\hat{\mu}^*\right) = \left(1-\frac{n}{N}\right)\frac{S^2}{n} + \frac{(k-1)}{n}W_{nr}S_{nr}^2$$

3. จากบทแทรก 1 จะได้ว่า s_{rr}^2 เป็นค่าประมาณของ S^2 และ s_{nr}^2 เป็นค่าประมาณของ S_{nr}^2 ดังนั้นจะได้ค่าประมาณความแปรปรวน เท่ากับ

$$v\left(\hat{\mu}^*\right) = \left(1-\frac{n}{N}\right)\frac{s_{rr}^2}{n} + \frac{(k-1)}{n}w_{nr}s_{nr}^2$$

เมื่อ

$$s_{rr}^2 = \frac{1}{n_{rr}-1}\left[\sum_{i=1}^{n_{rr}}(y_i-\hat{\mu}_{rr})^2\right] \text{ โดยที่ } \hat{\mu}_{rr} = \frac{1}{n_{rr}}(y_r+y'_{nr})$$

$$s_{nr}^2 = \frac{1}{n'_{nr}-1}\sum_{i=1}^{n'_{nr}}(y_i-\hat{\mu}'_{nr})^2 \text{ โดยที่ } \hat{\mu}'_{nr} = \frac{y'_{nr}}{n'_{nr}}$$

หากในกรอบตัวอย่างมีตัวแปรช่วย ที่มีความสัมพันธ์กับตัวแปรที่ศึกษา และคาดว่าตัวแปรช่วยดังกล่าวจะทำให้การประมาณค่ามีความแม่นยำมากขึ้น จากทฤษฎี 2 ทฤษฎี 3 และ

ทฤษฎี 9 จะได้ตัวประมาณอัตราส่วน และตัวประมาณการถดถอย เมื่อเกิดปัญหาการไม่ได้รับความร่วมมือ ดังนี้

ทฤษฎี 10 การประมาณค่าเฉลี่ยประชากรด้วยตัวประมาณอัตราส่วน เมื่อเกิดปัญหาการไม่ได้รับความร่วมมือ และกำหนดให้

S_d^2 แทน ความแปรปรวนประชากรของตัวประมาณอัตราส่วน

S_{dnr}^2 แทน ความแปรปรวนประชากรที่ไม่ได้รับความร่วมมือของตัวประมาณอัตราส่วน

s_{drr}^2 แทน ค่าประมาณความแปรปรวนประชากรของตัวประมาณอัตราส่วน

s_{dnr}^2 แทน ค่าประมาณความแปรปรวนประชากรที่ไม่ได้รับความร่วมมือของตัวประมาณอัตราส่วน

1. $\hat{\mu}_R^*$ เป็นตัวประมาณค่าเฉลี่ยที่เอนเอียงของค่าเฉลี่ยประชากร μ

$$\hat{\mu}_R^* = \hat{R}^* \mu_x \tag{47}$$

เมื่อ

$$\mu_x = \frac{1}{N}\sum_{i=1}^{N} x_i \text{ และ } \hat{R}^* = \frac{\hat{\mu}^*}{\hat{\mu}_x}$$

โดยที่

$$\hat{\mu}^* = \frac{1}{n}(y_r + k y'_{nr}) \text{ และ } \hat{\mu}_x = \frac{1}{n}\sum_{i=1}^{n} x_i$$

2. ความแปรปรวนของตัวประมาณค่าเฉลี่ย

$$V\left(\hat{\mu}_R^*\right) = \left(1 - \frac{n}{N}\right)\frac{S_d^2}{n} + \frac{(k-1)}{n} W_{nr} S_{dnr}^2 \tag{48}$$

เมื่อ

$$S_d^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - Rx_i)^2 \text{ และ } S_{dnr}^2 = \frac{1}{N_{nr}-1}\sum_{i=1}^{N_{nr}}(y_i - R_{nr}x_i)^2$$

โดยที่

$$R = \frac{\mu}{\mu_x},\ R_{nr} = \frac{\mu_{nr}}{\mu_{xnr}},\ \mu_{nr} = \frac{1}{N_{nr}}\sum_{i=1}^{N_{nr}} y_i \text{ และ } \mu_{xnr} = \frac{1}{N_{nr}}\sum_{i=1}^{N_{nr}} x_i$$

3. ค่าประมาณความแปรปรวนตัวประมาณค่าเฉลี่ย

$$v\left(\hat{\mu}_R^*\right) = \left(1-\frac{n}{N}\right)\frac{s_{drr}^2}{n} + \frac{(k-1)}{n} w_{nr} s_{dnr}^2 \tag{49}$$

เมื่อ

$$s_{drr}^2 = \frac{1}{n_{rr}-1}\sum_{i=1}^{n_{rr}}(y_i - \hat{R}_{rr}x_i)^2 \text{ และ } s_{dnr}^2 = \frac{1}{n'_{nr}-1}\sum_{i=1}^{n'_{nr}}(y_i - \hat{R}'_{nr}x_i)^2$$

$$\hat{R}_{rr} = \frac{\hat{\mu}_{rr}}{\hat{\mu}_{xrr}},\ \hat{\mu}_{rr} = \frac{1}{n_{rr}}\left(y_r + y'_{nr}\right) \text{ และ } \hat{\mu}_{xrr} = \frac{1}{n_{rr}}\left(x_r + x'_{nr}\right)$$

$$\hat{R}'_{nr} = \frac{\hat{\mu}'_{nr}}{\hat{\mu}'_{xnr}},\ \hat{\mu}'_{nr} \text{ และ } \hat{\mu}'_{xnr} = \frac{x'_{nr}}{n'_{nr}}$$

ทฤษฎี 11 การประมาณค่าเฉลี่ยประชากรด้วยตัวประมาณการถดถอย เมื่อเกิดปัญหาการไม่ได้รับความร่วมมือ และกำหนดให้

S_z^2 แทน ความแปรปรวนประชากรของตัวประมาณการถดถอย

S_{znr}^2 แทน ความแปรปรวนประชากรที่ไม่ให้ความร่วมมือของตัวประมาณการถดถอย

s_{zrr}^2 แทน ค่าประมาณความแปรปรวนประชากรของตัวประมาณการถดถอย

s_{znr}^2 แทน ค่าประมาณความแปรปรวนประชากรที่ไม่ให้ความร่วมมือของตัวประมาณการถดถอย

1. $\hat{\mu}_L^*$ เป็นค่าประมาณค่าเฉลี่ยที่เอนเอียงของค่าเฉลี่ยประชากร μ

$$\hat{\mu}_L^* = \hat{\mu}^* + \hat{\beta}^*\left(\mu_x - \hat{\mu}_x\right) \tag{50}$$

เมื่อ

$$\hat{\beta}^* = \frac{s_{xy}^*}{s_x^{*2}},\ s_{xy}^* = \frac{1}{n_{rr}-1}\sum_{i=1}^{n_{rr}}(x_i - \hat{\mu}_{xrr})(y_i - \hat{\mu}_{rr}) \text{ เมื่อ } \hat{\mu}_{rr} = \frac{1}{n_{rr}}\left(y_r + y'_{nr}\right)$$

$$s_x^{*2} = \frac{1}{n_{rr}-1}\sum_{i=1}^{n_{rr}}(x_i - \hat{\mu}_{xrr})^2 \text{ เมื่อ } \hat{\mu}_{xrr} = \frac{1}{n_{rr}}\left(x_r + x'_{nr}\right)$$

2. ความแปรปรวนของตัวประมาณค่าเฉลี่ย

$$V\left(\hat{\mu}_L^*\right) = \left(1 - \frac{n}{N}\right)\frac{S_z^2}{n} + \frac{(k-1)}{n}W_{nr}S_{znr}^2 \tag{51}$$

เมื่อ

$$S_{znr}^2 = \frac{1}{N_{nr}-1}\sum_{i=1}^{N_{nr}}\left[(y_i - \mu_{nr}) - \beta_{nr}(x_i - \mu_{xnr})\right]^2$$

$$\beta_{nr} = \frac{\sum_{i=1}^{N_{nr}}(x_i - \mu_{xnr})(y_i - \mu_{nr})}{\sum_{i=1}^{N_{nr}}(x_i - \mu_{xnr})^2},\ \mu_{nr} = \frac{1}{N_{nr}}\sum_{i=1}^{N_{nr}}y_i \text{ และ } \mu_{xnr} = \frac{1}{N_{nr}}\sum_{i=1}^{N_{nr}}x_i$$

3. ค่าประมาณความแปรปรวนตัวประมาณค่าเฉลี่ย

$$v\left(\hat{\mu}_L^*\right) = \left(1 - \frac{n}{N}\right)\frac{s_{zrr}^2}{n} + \frac{(k-1)}{n}w_{nr}s_{znr}^2 \tag{52}$$

เมื่อ

$$s_{zrr}^2 = \frac{1}{n_{rr}-2}\left[\sum_{i=1}^{n_{rr}}(y_i-\hat{\mu}_{rr})^2 - \hat{\beta}^{*2}\sum_{i=1}^{n_{rr}}(x_i-\hat{\mu}_{xrr})^2\right]$$

$$s_{znr}^2 = \frac{1}{n'_{nr}-2}\left(\sum_{i=1}^{n'_{nr}}(y_i-\hat{\mu}'_{nr})^2 - \hat{\beta}'^2_{nr}\sum_{i=1}^{n'_{nr}}(x_i-\hat{\mu}'_{xnr})^2\right)$$

$$\hat{\beta}'_{nr} = \frac{\sum_{i=1}^{n'_{nr}}(x_i-\hat{\mu}'_{xnr})(y_i-\hat{\mu}'_{nr})}{\sum_{i=1}^{n'_{nr}}(x_i-\hat{\mu}'_{xnr})^2},\ \hat{\mu}'_{nr} = \frac{y'_{nr}}{n'_{nr}}\ \text{และ}\ \hat{\mu}'_{xnr} = \frac{x'_{nr}}{n'_{nr}}$$

ตัวอย่าง 7 เพื่อให้เข้าใจง่ายสำหรับการสุ่มตัวอย่างเมื่อเกิดปัญหาการไม่ได้รับความร่วมมือ จากตัวอย่าง 1 จะได้กรอบตัวอย่างที่สมบูรณ์ (ดังแสดงในภาพ 4) ขนาด $N = N_T = 100$ และค่าเฉลี่ยประชากร $\mu = \mu_T = 5.0130$

1	2	3	4	5	6	7	8	9	10	11	12	13
(5, 5.6)	(3, 4.5)	(5, 6.3)	(5, 6.3)	(2, 3.3)	(3, 4.4)	(3, 4.6)	(5, 5.1)	(4, 5.4)	(4, 4.6)	(4, 5.2)	(3, 4.0)	(5, 6.0)
14	15	16	17	18	19	20	21	22	23	24	25	26
(4, 4.8)	(7, 6.3)	(3, 4.8)	(2, 3.3)	(7, 7.4)	(4, 5.8)	(4, 4.6)	(3, 4.3)	(3, 4.3)	(4, 5.2)	(5, 6.1)	(3, 4.4)	(6, 6.3)
27	28	29	30	31	32	33	34	35	36	37	38	39
(4, 5.5)	(5, 5.8)	(4, 5.3)	(4, 4.9)	(4, 5.3)	(5, 6.1)	(4, 4.4)	(3, 4.8)	(3, 3.9)	(2, 3.2)	(5, 6.0)	(3, 4.7)	(4, 5.0)
40	41	42	43	44	45	46	47	48	49	50	51	52
(5, 5.7)	(4, 5.5)	(2, 3.7)	(3, 4.0)	(2, 3.8)	(5, 5.5)	(4, 4.3)	(3, 3.7)	(3, 4.7)	(4, 4.7)	(2, 4.1)	(2, 4.1)	(4, 5.2)
53	54	55	56	57	58	59	60	61	62	63	64	65
(3, 4.3)	(4, 5.5)	(2, 3.5)	(6, 6.5)	(5, 6.1)	(5, 6.1)	(5, 5.5)	(3, 4.6)	(4, 4.8)	(3, 5.0)	(3, 4.3)	(2, 4.2)	(4, 5.3)
66	67	68	69	70	71	72	73	74	75	76	77	78
(5, 5.7)	(5, 6.0)	(3, 4.9)	(5, 5.5)	(6, 6.6)	(6, 6.8)	(4, 4.7)	(2, 3.3)	(4, 5.2)	(3, 4.4)	(6, 6.8)	(5, 4.9)	(1, 3.6)
79	80	81	82	83	84	85	86	87	88	89	90	91
(4, 4.4)	(4, 5.4)	(4, 5.5)	(5, 5.7)	(4, 5.5)	(4, 5.0)	(4, 4.8)	(4, 4.7)	(4, 4.8)	(6, 5.9)	(5, 5.5)	(2, 3.8)	(1, 2.6)
92	93	94	95	96	97	98	99	100				
(4, 5.4)	(5, 4.9)	(3 4.2)	(5, 5.7)	(6, 6.3)	(4, 4.2)	(4, 5.2)	(5, 5.8)	(5, 5.1)				

ภาพ 4 แสดงตัวอย่างสำหรับกรอบตัวอย่างที่สมบูรณ์

การสุ่มตัวอย่างแบบง่ายไม่แทนที่ขนาด n จากประชากรในกรอบตัวอย่าง N หน่วย เมื่อเกิดปัญหาการไม่ได้รับความร่วมมือ และมีการสุ่มตัวอย่างย่อยในกลุ่มตัวอย่างที่ไม่ให้ความร่วมมือ มีขั้นตอนดังนี้

ขั้นตอน 1 สุ่มตัวอย่างแบบง่ายไม่แทนที่ขนาด n จากประชากรในกรอบตัวอย่าง N หน่วย กำหนดให้ n_r เป็นขนาดตัวอย่างที่ให้ความร่วมมือ และ n_{nr} เป็นขนาดตัวอย่างที่ไม่ให้ความร่วมมือ สมมติว่าสุ่มตัวอย่างแบบง่ายไม่แทนที่ 30% ของประชากรในกรอบตัวอย่างขนาด $N = 100$ ดังนั้น $n = 30$ และสมมติว่าเกิดปัญหาการไม่ได้รับความร่วมมือ 30% ดังนั้นเราจะได้ตัวอย่างที่ให้ความร่วมมือ $n_r = 21$ หน่วย และตัวอย่างที่ไม่ให้ความร่วมมือ $n_{nr} = 9$ หน่วย ดังนี้

ตาราง 4 แสดงตัวอย่างข้อมูลของตัวอย่างที่สุ่มจากกรอบตัวอย่างสมบูรณ์ เมื่อเกิดปัญหาการไม่ได้รับความร่วมมือ

หน่วยที่ (i)	y_i	x_i	**หน่วยที่ (i)**	y_i	x_i
59	5.5	5	31	5.3	4
97	4.2	4	1	5.6	5
23	(5.2)	4	49	(4.7)	4
45	5.5	5	71	6.8	6
55	3.5	2	85	(4.8)	4
12	4.0	3	53	4.3	3
29	5.3	4	17	3.3	2
61	(4.8)	4	75	4.4	3
5	(3.3)	2	89	5.5	5
95	5.7	5	67	(6.0)	5
93	4.9	5	33	4.4	4
27	(5.5)	4	99	5.8	5
77	(4.9)	5	63	4.3	3
41	5.5	4	10	(4.6)	4
87	4.8	4	26	6.3	6
ผลรวมกลุ่มที่ให้ความร่วมมือ				$y_r = 104.9$	$x_r = 87$
ผลรวมกลุ่มที่ไม่ให้ความร่วมมือ				($y_{nr} = 43.8$)	$x_{nr} = 36$

หมายเหตุ: ตัวเลขที่อยู่ในวงเล็บ แทน ข้อมูลที่ไม่ได้รับความร่วมมือในการให้ข้อมูล

ขั้นตอน 2 สุ่มตัวอย่างย่อยในกลุ่มที่ไม่ให้ความร่วมมือขนาด $n'_{nr} = n_{nr} / k$ = 9/2 ≈ 5 (กำหนด k = 2) จากตัวอย่างที่ไม่ให้ความร่วมมือขนาด n_{nr} = 9 (ดังแสดงในตาราง 5)

จากตัวอย่างที่ให้ความร่วมมือในการสุ่มตัวอย่างครั้งที่ 1 ขนาด n_{1r} = 5 มีค่าสัมประสิทธิ์สหสัมพันธ์ เท่ากับ 0.923 และตัวอย่างที่ให้ความร่วมมือในการสุ่มตัวอย่างครั้งที่ 2 ขนาด n'_{nr} = 5 มีค่าสัมประสิทธิ์สหสัมพันธ์ เท่ากับ 0.875 และค่าสัมประสิทธิ์สหสัมพันธ์ของตัวอย่างที่ให้ความร่วมมือทั้งหมดขนาด n_{rr} = 21 + 5 = 26 มีค่าเท่ากับ 0.920

ตาราง 5 แสดงตัวอย่างข้อมูลที่ได้จากการสุ่มตัวอย่าง เมื่อเกิดปัญหาการไม่ได้รับความร่วมมือ

หน่วยที่ (i)	y_i	x_i
49	4.7	4
61	4.8	4
27	5.5	4
23	5.2	4
67	6.0	5
ผลรวม	y'_{nr} = 26.2	x'_{nr} = 21

ตัวอย่าง 7 จากทฤษฎี 9 จะได้ตัวประมาณอย่างง่ายของค่าเฉลี่ยประชากร μ มีค่าเท่ากับ

$$\hat{\mu}^* = \frac{1}{n}\left(y_r + k\,y'_{nr}\right) = \frac{1}{30}[104.9+1.8(26.2)] = 5.0687$$

โดยคำนวณค่าต่าง ๆ ได้ดังนี้

$$\hat{\mu}'_{nr} = \frac{y'_{nr}}{n'_{nr}} = \frac{26.2}{5} = 5.24$$

$$s^2_{nr} = \frac{1}{n'_{nr}-1}\sum_{i=1}^{n'_{nr}}(y_i - \hat{\mu}'_{nr})^2 = \frac{1}{5\text{-}1}(1.1320) = 0.2830$$

$$\hat{\mu}_{rr} = \frac{1}{n_{rr}}\left(y_r + y'_{nr}\right) = \frac{1}{21+5}(104.9+26.2) = \frac{1}{26}(131.1) = 5.0423$$

$$s_{rr}^2 = \frac{1}{n_{rr}-1}\sum_{i=1}^{n_{rr}}(y_i - \hat{\mu}_{rr})^2 = \frac{1}{26-1}(17.5035) = 0.7001$$

$$k = \frac{n_{nr}}{n'_{nr}} = \frac{9}{5} = 1.8 \text{ และ } w_{nr} = \frac{n_{nr}}{n} = \frac{9}{30} = 0.3$$

มีค่าเอนเอียง เท่ากับ

$$Bias(\hat{\mu}^*) = \hat{\mu}^* - \mu = 5.0687 - 5.0130 = 0.0557$$

ค่าประมาณความแปรปรวนตัวประมาณค่าเฉลี่ยประชากร มีค่าเท่ากับ

$$v\left(\hat{\mu}^*\right) = \left(1 - \frac{n}{N}\right)\frac{s_{rr}^2}{n} + \frac{(k-1)}{n} w_{nr} s_{nr}^2$$

$$= \left(1 - \frac{30}{100}\right)\frac{0.7001}{30} + \frac{(1.8-1)}{30}\frac{9}{30}(0.2830)$$

$$= 0.0186$$

และค่าความคลาดเคลื่อนกำลังสองเฉลี่ย มีค่าเท่ากับ

$$MSE(\hat{\mu}^*) = \left[Bias(\hat{\mu}^*)\right]^2 + v(\hat{\mu}^*) = (0.0557)^2 + 0.0186 = 0.0217$$

ตัวอย่าง 8 จากทฤษฎี 10 จะได้ตัวประมาณอัตราส่วนของค่าเฉลี่ยประชากร μ มีค่าเท่ากับ

$$\hat{\mu}_R^* = \hat{R}^*\mu_x = \frac{5.0687}{4.1000}(3.9600) = 4.8956$$

โดยคำนวณค่าต่าง ๆ ได้ดังนี้

$$\hat{\mu}_x = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{30}(87+36) = 4.10 \text{ และ } \mu_x = \mu_{Tx} = 3.96$$

$$\hat{R}^* = \frac{\hat{\mu}^*}{\hat{\mu}_x} = \frac{5.0687}{4.1000} = 1.2363$$

$$\hat{\mu}_{xrr} = \frac{1}{n_{rr}}(x_r + x'_{nr}) = \frac{1}{21+5}(87+21) = 4.1538$$

$$\hat{R}_{rr} = \frac{\hat{\mu}_{rr}}{\hat{\mu}_{xrr}} = \frac{5.0423}{4.1538} = 1.2139$$

$$\hat{\mu}'_{xnr} = \frac{x'_{nr}}{n'_{nr}} = \frac{21}{5} = 4.2$$

$$\hat{R}'_{nr} = \frac{\hat{\mu}'_{nr}}{\hat{\mu}'_{xnr}} = \frac{5.24}{4.2} = 1.2476$$

$$s_{dnr}^2 = \frac{1}{n'_{nr}-1}\sum_{i=1}^{n'_{nr}}(y_i - \hat{R}'_{nr}x_i)^2 = \frac{1}{5-1}(0.4809) = 0.1202$$

$$s_{drr}^2 = \frac{1}{n_{rr}-1}\sum_{i=1}^{n_{rr}}(y_i - \hat{R}_{rr}x_i)^2 = \frac{1}{26-1}(9.4679) = 0.3787$$

มีค่าเอนเอียง เท่ากับ

$$Bias(\hat{\mu}_R^*) = \hat{\mu}_R^* - \mu = 4.8956 - 5.0130 = -0.1174$$

ค่าประมาณความแปรปรวนตัวประมาณค่าเฉลี่ยประชากร มีค่าเท่ากับ

$$v\left(\hat{\mu}_R^*\right)=\left(1-\frac{n}{N}\right)\frac{s_{drr}^2}{n}+\frac{(k-1)}{n}w_{nr}s_{dnr}^2$$

$$=\left(1-\frac{30}{100}\right)\frac{0.1202}{30}+\frac{(1.8-1)}{30}\frac{9}{30}(0.3787)\ =0.0058$$

และค่าความคลาดเคลื่อนกำลังสองเฉลี่ย มีค่าเท่ากับ

$$MSE(\hat{\mu}_R^*)=\left[Bias(\hat{\mu}_R^*)\right]^2+v(\hat{\mu}_R^*)\ =(-0.1174)^2+0.0062\ =0.0200$$

ตัวอย่าง 9 จากทฤษฎี 11 จะได้ตัวประมาณการถดถอยของค่าเฉลี่ยประชากร μ มีค่าเท่ากับ

$$\hat{\mu}_L^*=\hat{\mu}^*+\hat{\beta}^*\left(\mu_x-\hat{\mu}_x\right)\ =5.0687+0.7278\ (3.96-4.10)=4.9668$$

โดยคำนวณค่าต่าง ๆ ได้ดังนี้

$$s_{xy}^*=\frac{1}{n_{rr}-1}\sum_{i=1}^{n_{rr}}(x_i-\hat{\mu}_{xrr})(y_i-\hat{\mu}_{rr})\ =\frac{1}{26-1}(19.9308)\ =0.7972$$

$$s_x^{*2}=\frac{1}{n_{rr}-1}\sum_{i=1}^{n_{rr}}(x_i-\hat{\mu}_{xrr})^2\ =\frac{1}{26-1}(27.3846)\ =1.0954$$

$$\hat{\beta}^*=\frac{s_{xy}^*}{s_x^{*2}}\ =\frac{0.7972}{1.0954}\ =0.7278$$

$$s_{zrr}^2=\frac{1}{n_{rr}-2}\left[\sum_{i=1}^{n_{rr}}(y_i-\hat{\mu}_{rr})^2-\hat{\beta}^{*2}\sum_{i=1}^{n_{rr}}(x_i-\hat{\mu}_{xrr})^2\right]$$

$$= \frac{1}{26\text{-}2}[17.5035\text{-}(0.7278)^2(27.3846)] = 0.1249$$

$$\hat{\beta}'_{nr} = \frac{\sum_{i=1}^{n'_{nr}}(x_i - \hat{\mu}'_{xnr})(y_i - \hat{\mu}'_{nr})}{\sum_{i=1}^{n'_{nr}}(x_i - \hat{\mu}'_{xnr})^2} = \frac{0.76}{0.80} = 0.95$$

$$s^2_{znr} = \frac{1}{n'_{nr} - 2}\left(\sum_{i=1}^{n'_{nr}}(y_i - \hat{\mu}'_{nr})^2 - \hat{\beta}'^2_{nr}\sum_{i=1}^{n'_{nr}}(x_i - \hat{\mu}'_{xnr})^2\right)$$

$$= \frac{1}{5\text{-}2}[1.1320\text{-}(0.95)^2(0.8)] = 0.1367$$

มีค่าเอนเอียง เท่ากับ

$$Bias(\hat{\mu}^*_L) = \hat{\mu}^*_L - \mu = 4.9668 - 5.0130 = -0.0462$$

ค่าประมาณความแปรปรวนตัวประมาณค่าเฉลี่ยประชากร มีค่าเท่ากับ

$$v\left(\hat{\mu}^*_L\right) = \left(1 - \frac{n}{N}\right)\frac{s^2_{zrr}}{n} + \frac{(k-1)}{n}w_{nr}s^2_{znr}$$

$$= \left(1\text{-}\frac{30}{100}\right)\frac{0.1249}{30} + \frac{(1.8\text{-}1)}{30}\frac{9}{30}(0.1367) = 0.0040$$

และค่าความคลาดเคลื่อนกำลังสองเฉลี่ย มีค่าเท่ากับ

$$MSE(\hat{\mu}^*_L) = \left[Bias(\hat{\mu}^*_L)\right]^2 + v(\hat{\mu}^*_L) = (-0.0462)^2 + 0.0040 = 0.0061$$

จากตัวอย่าง 7-9 จะได้

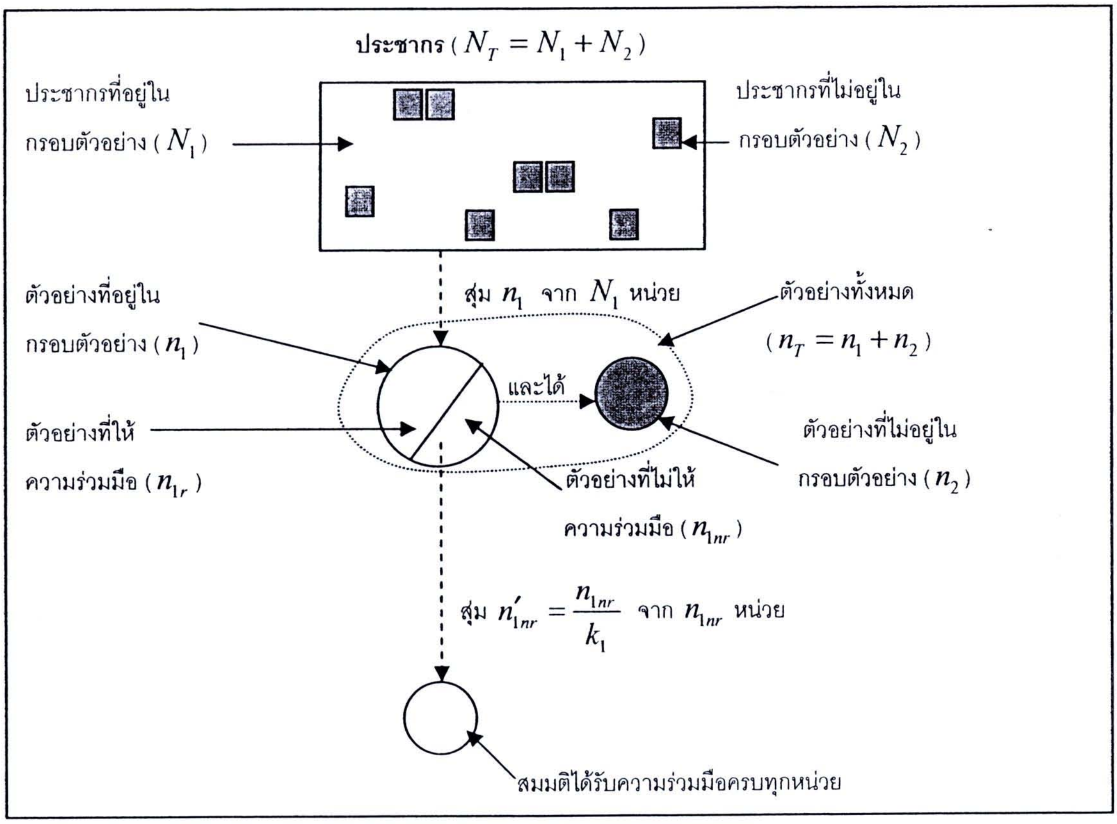
$$MSE(\hat{\mu}_L^*) = 0.0061 < MSE(\hat{\mu}_R^*) = 0.0200 < MSE(\hat{\mu}^*) = 0.0217$$

การประมาณค่าเฉลี่ยประชากร เมื่อเกิดปัญหากรอบตัวอย่างไม่ครอบคลุมและการไม่ได้รับความร่วมมือ

ความคลาดเคลื่อนในการสำรวจด้วยตัวอย่างที่เกิดจากการไม่ได้สังเกตค่าตัวแปรที่ศึกษา ประกอบด้วย 1) ความคลาดเคลื่อนจากการไม่ครอบคลุมหรือความความคลาดเคลื่อนจากกรอบตัวอย่าง และ 2) ความเคลื่อนจากการไม่ได้รับความร่วมมือ ซึ่งคลาดเคลื่อนทั้ง 2 ประการ ต่างก็ทำให้การประมวลผล และการวิเคราะห์ข้อมูลมีความเอนเอียงไปจากค่าที่แท้จริงของประชากร หากความคลาดเคลื่อนจากกรอบตัวอย่างไม่ครอบคลุม และการไม่ได้รับความร่วมมือเกิดขึ้นพร้อมกัน ยิ่งส่งผลต่อความเอนเอียงอย่างรุนแรงในการประมาณค่าที่แท้จริงของประชากร (Särndal and Lundström, 2005, pp. 179-181) เนื่องจากการเพิ่มขนาดตัวอย่างจะทำให้ความเอนเอียงเนื่องจากกรอบตัวอย่างไม่สมบูรณ์มีค่ามากขึ้น และเมื่ออัตราการไม่ได้รับความร่วมมือมากขึ้น ความเอนเอียงก็จะมีค่าสูงขึ้นเช่นเดียวกัน ดังนั้นความคลาดเคลื่อนที่เกิดจากกรอบตัวอย่างและการไม่ได้รับความร่วมมือจึงมักจะได้รับการพิจารณาร่วมกัน

ข้อตกลงเบื้องต้น การวิจัยครั้งนี้ศึกษาปัญหาการไม่ได้รับความร่วมมือบางหน่วยในตัวอย่างที่อยู่ในกรอบตัวอย่าง และศึกษากรณีที่กรอบตัวอย่างที่ใช้ในการสำรวจนั้นมีการระบุตำแหน่งที่ตั้งของหน่วยตัวอย่างที่ชัดเจน จึงทำให้สามารถหาที่อยู่หรือหน่วยตัวอย่างที่อยู่ในกรอบตัวอย่าง และหาหน่วยตัวอย่างที่ไม่อยู่ในกรอบตัวอย่างได้ ดังนั้นการไม่ได้รับความร่วมมือบางหน่วย อาจเกิดขึ้นจากการไม่สามารถรวบรวมค่าของตัวแปรจากบางหน่วยตัวอย่างด้วยเหตุผลต่าง ๆ เช่น การปฏิเสธที่จะให้ความร่วมมือ แบบสอบถามสูญหายภายหลังการสัมภาษณ์ เป็นต้น

การประมาณค่าเฉลี่ยประชากร μ_T จะทำการสุ่มตัวอย่างขนาด n_1 จากประชากรในกรอบตัวอย่าง N_1 หน่วย และในระหว่างการสำรวจตัวอย่างหน่วยที่ i จะเก็บรวบรวมข้อมูลว่ามีหน่วยตัวอย่างใดบ้างที่ไม่อยู่ในกรอบตัวอย่างระหว่างตัวอย่างที่ i และหน่วยที่ $i+1$ พร้อมทั้งเก็บข้อมูลที่สนใจศึกษาในตัวอย่างที่ไม่อยู่ในกรอบตัวอย่างในระหว่างการสุ่มตัวอย่างขนาด n_1 หน่วย ดังนั้นเราจะได้ตัวอย่างที่อยู่ในกรอบตัวอย่าง n_1 หน่วย และได้ตัวอย่างที่ไม่อยู่ในกรอบตัวอย่างขนาด n_2 ซึ่งจะได้ตัวอย่างทั้งหมด n_T หน่วย แต่ในระหว่างการสำรวจพบว่าเกิดปัญหาการไม่ได้รับความร่วมมือในกลุ่มตัวอย่างที่สุ่มจากกรอบตัวอย่างขนาด n_1 กำหนดให้ n_{1r} เป็นขนาดตัวอย่างที่ให้ความร่วมมือ และ n_{1nr} เป็นขนาดตัวอย่างที่ไม่ให้ความร่วมมือ จากนั้นจึงสุ่มตัวอย่างย่อยในกลุ่มที่ไม่ความร่วมมือขนาด $n'_{1nr} = n_{1nr} / k_1$ (ดังแสดงในภาพ 5) เมื่อ k_1 เป็นค่าคงที่ที่กำหนดไว้ล่วงหน้า ($k_1 > 1$)



ภาพ 5 แสดงการสุ่มตัวอย่างเมื่อเกิดปัญหากรอบตัวอย่างไม่ครอบคลุม และการไม่ได้รับความร่วมมือ

กำหนดให้

N_{1nr} แทน จำนวนประชากรที่อยู่ในกรอบตัวอย่างที่ไม่ให้ความร่วมมือ

W_{1nr} แทน สัดส่วนประชากรที่อยู่ในกรอบตัวอย่าง และไม่ให้ความร่วมมือ ซึ่ง $W_{1nr} = N_{1nr} / N_1$

n_{1rr} แทน ขนาดตัวอย่างที่อยู่ในกรอบตัวอย่าง และให้ความร่วมมือทั้งหมด ซึ่ง $n_{1rr} = n_{1r} + n'_{1nr}$

y_{1r} แทน ผลรวมของตัวแปรที่ศึกษาขนาด n_{1r} ; $y_{1r} = \sum_{i=1}^{n_{1r}} y_{1i}$

y_{1nr} แทน ผลรวมของตัวแปรที่ศึกษาขนาด n_{1nr} ; $y_{1nr} = \sum_{i=1}^{n_{1nr}} y_{1i}$

y'_{1nr} แทน ผลรวมของตัวแปรที่ศึกษาขนาด n'_{1nr} ; $y'_{1nr} = \sum_{i=1}^{n'_{1nr}} y_{1i}$

x'_{1nr} แทน ผลรวมของตัวแปรช่วยขนาด n'_{1nr} ; $x'_{1nr} = \sum_{i=1}^{n'_{1nr}} x_{1i}$

y_1 แทน ผลรวมของตัวแปรที่ศึกษาขนาด n_1 ; $y_1 = y_{1r} + y_{1nr} = \sum_{i=1}^{n_1} y_{1i}$

x_1 แทน ผลรวมของตัวแปรช่วยขนาด n_1 ; $x_1 = \sum_{i=1}^{n_1} x_{1i}$

y_2 แทน ผลรวมของตัวแปรที่ศึกษาขนาด n_2 ; $y_2 = \sum_{i=1}^{n_2} y_{2i}$

x_2 แทน ผลรวมของตัวแปรช่วยขนาด n_2 ; $x_2 = \sum_{i=1}^{n_2} x_{2i}$

S^2_{1nr} แทน ความแปรปรวนประชากรที่อยู่ในกรอบตัวอย่าง และไม่ให้ความร่วมมือ ขนาด N_{1nr}

การสุ่มตัวอย่างแบบง่ายไม่แทนที่ การสุ่มตัวอย่างแบบง่ายไม่แทนที่ขนาด n_1 จากประชากรในกรอบตัวอย่างไม่ครอบคลุม N_1 หน่วย พบว่า มีตัวอย่างที่ไม่อยู่ในกรอบตัวอย่าง n_2 หน่วย (ได้รับความร่วมมือครบทุกหน่วย) และเกิดปัญหาการไม่ได้รับความร่วมมือในกลุ่มตัวอย่างที่อยู่ในกรอบตัวอย่าง ขนาด n_{1nr} และกลุ่มตัวอย่างที่อยู่ในกรอบตัวอย่างให้ความร่วมมือ $n_{1r} = n_1 - n_{1nr}$ หน่วย จึงสุ่มตัวอย่างย่อยในกลุ่มตัวอย่างที่ไม่ให้ความร่วมมือขนาด $n'_{1nr} = n_{1nr} / k_1$ ดังนั้นการประมาณค่าเฉลี่ยประชากร μ_T มีรายละเอียดดังนี้

ทฤษฎี 12 การประมาณค่าเฉลี่ยประชากรด้วยตัวประมาณอย่างง่าย เมื่อเกิดปัญหากรอบตัวอย่างไม่ครอบคลุมและการไม่ได้รับความร่วมมือ

1. $\hat{\mu}^*_{pst}$ เป็นตัวประมาณค่าเฉลี่ยที่เอนเอียงแบบไม่มีเงื่อนไขของค่าเฉลี่ยประชากร μ_T

$$\hat{\mu}^*_{pst} = \hat{W}_1 \hat{\mu}^*_1 + \hat{W}_2 \hat{\mu}_2 \tag{53}$$

เมื่อ

$$\hat{\mu}^*_1 = \frac{1}{n_1}(y_{1r} + k_1 y'_{1nr}) \text{ และ } \hat{\mu}_2 = \frac{1}{n_2}\sum_{i=1}^{n_2} y_{2i}$$

2. ความแปรปรวนแบบไม่มีเงื่อนไขของตัวประมาณค่าเฉลี่ยประชากร

$$V\left(\hat{\mu}^{*}_{pst}\right) \simeq \frac{1}{N_T^2}\left[W_1^2\mu_1^2+\left(W_2-1\right)^2\mu_2^2\right]V\left(\hat{N}_2\right)+\hat{W}_1^2\frac{\left(k_1-1\right)}{n_1}W_{1nr}S_{1nr}^2$$

$$+\left[\left(1-\frac{n_T}{\hat{N}_T}\right)\frac{1}{n_T}\sum_{h=1}^{2}\hat{W}_h S_h^2+\frac{1}{n_T^2}\sum_{h=1}^{2}\left(1-\hat{W}_h\right)S_h^2\right] \tag{54}$$

เมื่อ

$$S_h^2=\frac{1}{N_h-1}\sum_{i=1}^{N_h}(y_{hi}-\mu_h)^2 \text{ โดยที่ } \mu_h=\frac{1}{N_h}\sum_{i=1}^{N_h}y_{hi}$$

$$S_m^2=\frac{1}{N_1-1}\sum_{i=1}^{N_1}(m_{1i}-\bar{N}_2)^2 \text{ โดยที่ } \bar{N}_2=\frac{1}{N_1}\sum_{i=1}^{N_1}m_{1i}$$

$$S_{1nr}^2=\frac{1}{N_{1nr}-1}\sum_{i=1}^{N_{1nr}}(y_{1i}-\mu_{1nr})^2 \text{ โดยที่ } \mu_{1nr}=\frac{1}{N_{1nr}}\sum_{i=1}^{N_{1nr}}y_{1i}$$

3. ค่าประมาณความแปรปรวนแบบไม่มีเงื่อนไขของตัวประมาณค่าเฉลี่ยประชากร

$$v\left(\hat{\mu}^{*}_{pst}\right) \simeq \frac{1}{\hat{N}_T^2}\left[\hat{W}_1^2\hat{\mu}_1^{*2}+\left(\hat{W}_2-1\right)^2\hat{\mu}_2^{*2}\right]v\left(\hat{N}_2\right)+\hat{W}_1^2\frac{\left(k_1-1\right)}{n_1}w_{1nr}s_{1nr}^2$$

$$+\left[\left(1-\frac{n_T}{\hat{N}_T}\right)\frac{1}{n_T}\sum_{h=1}^{2}\hat{W}_h s_h^{*2}+\frac{1}{n_T^2}\sum_{h=1}^{2}\left(1-\hat{W}_h\right)s_h^{*2}\right] \tag{55}$$

เมื่อ

$$s_1^{*2}=\frac{1}{n_{1rr}-1}\sum_{i=1}^{n_{1rr}}(y_i-\hat{\mu}_{1rr})^2 \text{ โดยที่ } \hat{\mu}_{1rr}=\frac{1}{n_{1rr}}(y_{1r}+y'_{1nr})$$

$$s_2^2 = \frac{1}{n_2-1}\sum_{i=1}^{n_2}(y_{2i}-\hat{\mu}_2)^2 \text{ และ } s_{1nr}^2 = \frac{1}{n'_{1nr}-1}\sum_{i=1}^{n'_{1nr}}(y_{1i}-\hat{\mu}'_{1nr})^2 \text{ โดยที่ } \hat{\mu}'_{1nr} = \frac{y'_{1nr}}{n'_{1nr}}$$

พิสูจน์ การประมาณค่าเฉลี่ยประชากร ด้วยตัวประมาณอย่างง่าย เมื่อเกิดปัญหากรอบตัวอย่างไม่ครอบคลุม และการไม่ได้รับความร่วมมือ ดังนี้

$$\hat{\mu}^*_{pst} = \hat{W}_1\hat{\mu}^*_1 + \hat{W}_2\hat{\mu}^*_2$$

ค่าคาดหวังแบบไม่มีเงื่อนไขของตัวประมาณ $\hat{\mu}^*_{pst}$ ในการสุ่มตัวอย่าง มีค่าเท่ากับ

$$E\left(\hat{\mu}^*_{pst}\right) = E_1E_2\left[E_3\left(\hat{\mu}^*_{pst}\right) | n_1, n_2\right]$$

เมื่อ

E_1 คือ ค่าคาดหวังของพารามิเตอร์ μ_T สำหรับตัวอย่าง ขนาด n_T

E_2 คือ ค่าคาดหวังของ $E_3\left(\hat{\mu}^*_{pst}\right)$ เมื่อ n_1 และ n_2 เป็นค่าคงที่

E_3 คือ ค่าคาดหวังของพารามิเตอร์ μ_T สำหรับการสุ่มตัวอย่างย่อยในกลุ่มที่ไม่ให้ความร่วมมือ ขนาด n'_{1nr}

และค่าความแปรปรวนแบบไม่มีเงื่อนไขของตัวประมาณ $\hat{\mu}^*_{pst}$ ในการสุ่มตัวอย่าง มีค่าเท่ากับ

$$V\left(\hat{\mu}^*_{pst}\right) = V_1E_2\left[E_3\left(\hat{\mu}^*_{pst}\right) | n_1, n_2\right] + E_1V_2\left[E_3\left(\hat{\mu}^*_{pst}\right) | n_1, n_2\right] + E_1E_2V_3\left(\hat{\mu}^*_{pst}\right)$$

เมื่อ

V_1 คือ ค่าความแปรปรวนของพารามิเตอร์ μ_T สำหรับตัวอย่าง ขนาด n_T

V_2 คือ ค่าความแปรปรวนของ $E_3\left(\hat{\mu}^*_{pst}\right)$ เมื่อ n_1 และ n_2 เป็นค่าคงที่

V_3 คือ ค่าความแปรปรวนของพารามิเตอร์ μ_T สำหรับการสุ่มตัวอย่างย่อยในกลุ่มที่ไม่ให้ความร่วมมือ ขนาด n'_{1nr}

1.

$$E\left(\hat{\mu}^*_{pst}\right) = E_1E_2\left[E_3\left(\hat{\mu}^*_{pst}\right) \mid n_1, n_2\right]$$

$$= E_1E_2\left\{E_3\left[\left(\hat{W}_1\hat{\mu}^*_1 + \hat{W}_2\hat{\mu}^*_2\right) \mid n_1, n_2\right]\right\}$$

$$= E_1E_2\left\{E_3\left[\left(\hat{W}_1\frac{1}{n_1}(y_{1r} + \frac{n_{1nr}}{n'_{1nr}}y'_{1nr}) + \hat{W}_2\hat{\mu}_2\right) \mid n_1, n_2\right]\right\} \; ; \; \hat{\mu}_2 = \hat{\mu}^*_2$$

$$= E_1E_2\left[\left(\hat{W}_1\frac{1}{n_1}(y_{1r} + n_{1nr}E_3(\hat{\mu}'_{1nr})) + \hat{W}_2\hat{\mu}_2\right) \mid n_1, n_2\right] \; ; \; \hat{\mu}'_{1nr} = \frac{y'_{1nr}}{n'_{1nr}}$$

$$= E_1E_2\left[\left(\hat{W}_1\frac{1}{n_1}(y_{1r} + n_{1nr}\hat{\mu}_{1nr}) + \hat{W}_2\hat{\mu}_2\right) \mid n_1, n_2\right]$$

$$= E_1E_2\left[\left(\hat{W}_1\frac{1}{n_1}(y_{1r} + y_{1nr}) + \hat{W}_2\hat{\mu}_2\right) \mid n_1, n_2\right]$$

$$= E_1E_2\left[\left(\hat{W}_1\frac{1}{n_1}\sum_{i=1}^{n_1}y_i + \hat{W}_2\hat{\mu}_2\right) \mid n_1, n_2\right]$$

$$= E_1E_2\left[\left(\hat{W}_1\hat{\mu}_1 + \hat{W}_2\hat{\mu}_2\right) \mid n_1, n_2\right] \quad ; \; \hat{\mu}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1}y_{1i}$$

$$= E_1\left[E_2\left(\hat{\mu}_{pst} \mid n_1, n_2\right)\right] \quad ; \; \hat{\mu}_{pst} = \hat{W}_1\hat{\mu}_1 + \hat{W}_2\hat{\mu}_2$$

$$= E_1\left[\hat{W}_1\mu_1 + \hat{W}_2\mu_2\right]$$

$$\neq \mu_T \tag{56}$$

เนื่องจาก $E\left[\hat{\mu}_{pst} \mid n_1, n_2\right] = \hat{W}_1\mu_1 + \hat{W}_2\mu_2$ (จากทฤษฎี 6)

2. จากสมการ (23), (26) และ (32) ในทฤษฎี 6 ดังนี้

$$V\left[\hat{\mu}_{pst} \mid n_1, n_2\right] = \sum_{h=1}^{2} \hat{W}_h^2 \left(1 - \frac{n_h}{N_h}\right) \frac{S_h^2}{n_h}$$

$$E_1\left[V_2\left(\hat{\mu}_{pst} \mid n_1, n_2\right)\right] \simeq \left(1 - \frac{n_T}{\hat{N}_T}\right) \frac{1}{n_T} \sum_{h=1}^{2} \hat{W}_h S_h^2 + \frac{1}{n_T^2} \sum_{h=1}^{2} \left(1 - \hat{W}_h\right) S_h^2$$

และ

$$V_1\left[E_2\left(\hat{\mu}_{pst} \mid n_1, n_2\right)\right] = \frac{1}{N_T^2}\left[W_1^2\mu_1^2 + \left(W_2 - 1\right)^2 \mu_2^2\right] V\left(\hat{N}_2\right)$$

จะพิจารณา

$$V\left(\hat{\mu}_{pst}^*\right) = V_1E_2\left[E_3\left(\hat{\mu}_{pst}^*\right) \mid n_1, n_2\right] + E_1V_2\left[E_3\left(\hat{\mu}_{pst}^*\right) \mid n_1, n_2\right] + E_1E_2V_3\left(\hat{\mu}_{pst}^*\right)$$

ดังนี้

$$V_1E_2\left[E_3\left(\hat{\mu}_{pst}^*\right) \mid n_1, n_2\right] = V_1\left[E_2\left(\hat{\mu}_{pst} \mid n_1, n_2\right)\right]$$

$$= \frac{1}{N_T^2}\left[W_1^2\mu_1^2 + \left(W_2 - 1\right)^2 \mu_2^2\right] V\left(\hat{N}_2\right)$$

$$E_1V_2\left[E_3\left(\hat{\mu}_{pst}^*\right) \mid n_1, n_2\right] = E_1\left[V_2\left(\hat{\mu}_{pst} \mid n_1, n_2\right)\right]$$

$$\simeq \left(1 - \frac{n_T}{\hat{N}_T}\right) \frac{1}{n_T} \sum_{h=1}^{2} \hat{W}_h S_h^2 + \frac{1}{n_T^2} \sum_{h=1}^{2} \left(1 - \hat{W}_h\right) S_h^2$$

และ

$$E_1E_2V_3\left(\hat{\mu}^*_{pst}\right) = E_1E_2\left[\hat{W}_1^2\frac{1}{n_1^2}n_{1nr}^2V_3(\hat{\mu}'_{1nr})\right]$$

$$= E_1E_2\left[\hat{W}_1^2\frac{1}{n_1^2}n_{1nr}^2\left(1-\frac{n'_{1nr}}{n_{1nr}}\right)\frac{S_{1nr}^2}{n'_{1nr}}\right]$$

$$= E_1E_2\left[\hat{W}_1^2\frac{1}{n_1}\frac{n_{1nr}}{n_1}n_{1nr}\left(\frac{n_{1nr}-n'_{1nr}}{n_{1nr}}\right)\frac{S_{1nr}^2}{n'_{1nr}}\right]$$

$$= E_1E_2\left[\hat{W}_1^2\frac{1}{n_1}w_{1nr}\left(\frac{n_{1nr}}{n'_{1nr}}-\frac{n'_{1nr}}{n'_{1nr}}\right)S_{1nr}^2\right]$$

$$= E_1E_2\left[\hat{W}_1^2\frac{1}{n_1}(k_1-1)\,w_{1nr}S_{1nr}^2\right]$$

$$= E_1\left[\hat{W}_1^2\frac{(k_1-1)}{n_1}W_{1nr}S_{1nr}^2\right]$$

$$\approx \hat{W}_1^2\frac{(k_1-1)}{n_1}W_{1nr}S_{1nr}^2$$

ดังนั้นจะได้ความแปรปรวนของตัวประมาณ เท่ากับ

$$V\left(\hat{\mu}^*_{pst}\right) \simeq \frac{1}{N_T^2}\left[W_1^2\mu_1^2+\left(W_2-1\right)^2\mu_2^2\right]V\left(\hat{N}_2\right)+\hat{W}_1^2\frac{(k_1-1)}{n_1}W_{1nr}S_{1nr}^2$$

$$+\left[\left(1-\frac{n_T}{\hat{N}_T}\right)\frac{1}{n_T}\sum_{h=1}^{2}\hat{W}_hS_h^2+\frac{1}{n_T^2}\sum_{h=1}^{2}\left(1-\hat{W}_h\right)S_h^2\right]$$

เมื่อ

$$S_h^2 = \frac{1}{N_h - 1}\sum_{i=1}^{N_h}(y_{hi} - \mu_h)^2 \text{ โดยที่ } \mu_h = \frac{1}{N_h}\sum_{i=1}^{N_h} y_{hi}$$

$$S_{1nr}^2 = \frac{1}{N_{1nr} - 1}\sum_{i=1}^{N_{1nr}}(y_{1i} - \mu_{1nr})^2 \text{ โดยที่ } \mu_{1nr} = \frac{1}{N_{1nr}}\sum_{i=1}^{N_{1nr}} y_{1i}$$

3. ทำนองเดียวกันกับทฤษฎี 9 จะได้ว่า s_1^{*2} เป็นค่าประมาณความแปรปรวน S_1^2 และ s_{1nr}^2 เป็นค่าประมาณความแปรปรวน S_{1nr}^2 ดังนั้นค่าประมาณความแปรปรวนแบบไม่มีเงื่อนไขของตัวประมาณค่าเฉลี่ย มีค่าเท่ากับ

$$v\left(\hat{\mu}_{pst}^{*}\right) \simeq \frac{1}{\hat{N}_T^2}\left[\hat{W}_1^2\hat{\mu}_1^{*2} + \left(\hat{W}_2 - 1\right)^2\hat{\mu}_2^{*2}\right]v\left(\hat{N}_2\right) + \hat{W}_1^2\frac{(k_1 - 1)}{n_1}w_{1nr}s_{1nr}^2$$

$$+\left[\left(1 - \frac{n_T}{\hat{N}_T}\right)\frac{1}{n_T}\sum_{h=1}^{2}\hat{W}_h s_h^{*2} + \frac{1}{n_T^2}\sum_{h=1}^{2}\left(1 - \hat{W}_h\right)s_h^{*2}\right]$$

เมื่อ

$$v\left(\hat{N}_2\right) = N_1^2\left(1 - \frac{n_1}{N_1}\right)\frac{s_m^2}{n_1} \text{ (จากทฤษฎี 4)}$$

$$s_1^{*2} = \frac{1}{n_{1rr} - 1}\sum_{i=1}^{n_{1rr}}(y_{1i} - \hat{\mu}_{1rr})^2 \text{ โดยที่ } \hat{\mu}_{1rr} = \frac{1}{n_{1rr}}\left(y_{1r} + y'_{1nr}\right)$$

$$s_{1nr}^2 = \frac{1}{n'_{1nr} - 1}\sum_{i=1}^{n'_{1nr}}(y_{1i} - \hat{\mu}'_{1nr})^2 \text{ โดยที่ } \hat{\mu}'_{1nr} = \frac{y'_{1nr}}{n'_{1nr}}$$

$$s_2^2 = \frac{1}{n_2 - 1}\sum_{i=1}^{n_2}(y_{2i} - \hat{\mu}_2)^2$$

จากทฤษฎี 12 การประมาณค่าเฉลี่ยเมื่อเกิดปัญหากรอบตัวอย่างไม่ครอบคลุม และการไม่ได้รับความร่วมมือ เมื่อพบว่าในกรอบตัวอย่างมีตัวแปรช่วยที่มีความสัมพันธ์กับตัวแปรที่ศึกษาที่ช่วยให้การประมาณค่ามีความแม่นยำมากขึ้น จึงนำประยุกต์ใช้กับการประมาณค่าเมื่อเกิดปัญหากรอบตัวอย่างไม่ครอบคลุมในทฤษฎี 7 และทฤษฎี 8 และการประมาณค่า เมื่อเกิดปัญหาการไม่ได้รับความร่วมมือในทฤษฎี 11 และทฤษฎี 12 ได้ตัวประมาณอัตราส่วน และตัวประมาณการถดถอย เมื่อเกิดปัญหากรอบตัวอย่างไม่ครอบคลุม และการไม่ได้รับความร่วมมือ ดังนี้

ทฤษฎี 13 การประมาณค่าเฉลี่ยประชากรด้วยตัวประมาณอัตราส่วน เมื่อเกิดปัญหากรอบตัวอย่างไม่ครอบคลุมและการไม่ได้รับความร่วมมือ

1. $\hat{\mu}^*_{Rpst}$ เป็นตัวประมาณค่าเฉลี่ยที่เอนเอียงแบบไม่มีเงื่อนไขของค่าเฉลี่ยประชากร μ_T

$$\hat{\mu}^*_{Rpst} = \hat{W}_1\hat{\mu}^*_{R1} + \hat{W}_2\hat{\mu}^*_{R2} \tag{57}$$

เมื่อ

$$\hat{\mu}^*_{R1} = \frac{\hat{\mu}^*_1}{\hat{\mu}_{x1}}\mu_{x1} = \hat{R}^*_1\mu_{x1} \text{ และ } \hat{\mu}^*_{R2} \approx \hat{R}_2\mu_{x1}$$

$$\hat{\mu}^*_1 = \frac{1}{n_1}(y_{1r} + k_1 y'_{1nr}) \text{ และ } \hat{\mu}^*_2 = \hat{\mu}_2 = \frac{1}{n_2}\sum_{i=1}^{n_2} y_{2i}$$

2. ความแปรปรวนแบบไม่มีเงื่อนไขของตัวประมาณค่าเฉลี่ย

$$V\left(\hat{\mu}^*_{Rpst}\right) \simeq \frac{1}{N_T^2}\left[W_1^2\mu_1^2 + \left(W_2 - 1\right)^2\mu_2^2\right]V\left(\hat{N}_2\right) + \hat{W}_1^2\frac{(k_1-1)}{n_1}W_{1nr}S^2_{d1nr}$$

$$+\left[\left(1-\frac{n_T}{\hat{N}_T}\right)\frac{1}{n_T}\sum_{h=1}^{2}\hat{W}_h S^2_{dh} + \frac{1}{n_T^2}\sum_{h=1}^{2}\left(1-\hat{W}_h\right)S^2_{dh}\right] \tag{58}$$

เมื่อ

$$S_{dh}^2 = \frac{1}{N_h - 1}\sum_{i=1}^{N_h}(y_{hi} - R_h x_{hi})^2 \text{ โดยที่ } R_h = \frac{\mu_h}{\mu_{xh}}$$

$$S_{d1nr}^2 = \frac{1}{N_{1nr} - 1}\sum_{i=1}^{N_{1nr}}(y_{1i} - R_{1nr} x_{1i})^2 \text{ โดยที่ } R_{1nr} = \frac{\mu_{1nr}}{\mu_{x1nr}}$$

$$\mu_{1nr} = \frac{1}{N_{1nr}}\sum_{i=1}^{N_{1nr}} y_{1i} \text{ และ } \mu_{x1nr} = \frac{1}{N_{1nr}}\sum_{i=1}^{N_{1nr}} x_{1i}$$

3. ค่าประมาณความแปรปรวนแบบไม่มีเงื่อนไขของตัวประมาณค่าเฉลี่ย

$$v\left(\hat{\mu}_{Rpst}^*\right) \simeq \frac{1}{\hat{N}_T^2}\left[\hat{W}_1^2\hat{\mu}_{R1}^{*2} + \left(\hat{W}_2 - 1\right)^2 \hat{\mu}_{R2}^{*2}\right] v\left(\hat{N}_2\right) + \hat{W}_1^2 \frac{(k_1 - 1)}{n_1} w_{1nr} s_{d1nr}^2$$

$$+ \left[\left(1 - \frac{n_T}{\hat{N}_T}\right)\frac{1}{n_T}\sum_{h=1}^{2}\hat{W}_h s_{dh}^{*2} + \frac{1}{n_T^2}\sum_{h=1}^{2}\left(1 - \hat{W}_h\right) s_{dh}^{*2}\right] \tag{59}$$

เมื่อ

$$s_{d1}^{*2} = \frac{1}{n_{1rr} - 1}\sum_{i=1}^{n_{1rr}}(y_{1i} - \hat{R}_{1rr} x_{1i})^2 \text{ โดยที่ } \hat{R}_{1rr} = \frac{\hat{\mu}_{1rr}}{\hat{\mu}_{x1rr}}$$

$$\hat{\mu}_{1rr} = \frac{1}{n_{1rr}}\left(y_{1r} + y'_{1nr}\right) \text{ และ } \hat{\mu}_{x1rr} = \frac{1}{n_{1rr}}\left(x_{1r} + x'_{1nr}\right)$$

$$s_{d1nr}^2 = \frac{1}{n'_{1nr} - 1}\sum_{i=1}^{n'_{1nr}}(y_i - \hat{R}'_{1nr} x_{1i})^2 \text{ โดยที่ } \hat{R}'_{1nr} = \frac{\hat{\mu}'_{1nr}}{\hat{\mu}'_{x1nr}}$$

$$\hat{\mu}'_{1nr} = \frac{y'_{1nr}}{n'_{1nr}} \text{ และ } \hat{\mu}'_{x1nr} = \frac{x'_{1nr}}{n'_{1nr}}$$

$$s_{d2}^{*2} = s_{d2}^{2} = \frac{1}{n_2 - 1}\sum_{i=1}^{n_2}(y_i - \hat{R}_2 x_{2i})^2 \text{ โดยที่ } \hat{R}_2 = \frac{\hat{\mu}_2}{\hat{\mu}_{x2}}$$

ทฤษฎี 14 การประมาณค่าเฉลี่ยประชากรด้วยตัวประมาณการถดถอย เมื่อเกิดปัญหากรอบตัวอย่างไม่ครอบคลุมและการไม่ได้รับความร่วมมือ

1. $\hat{\mu}^*_{Lpst}$ เป็นตัวประมาณค่าเฉลี่ยที่เอนเอียงแบบไม่มีเงื่อนไขของค่าเฉลี่ยประชากร μ_T

$$\hat{\mu}^*_{Lpst} = \hat{W}_1\hat{\mu}^*_{R1} + \hat{W}_2\hat{\mu}^*_{R2} \tag{60}$$

เมื่อ

$$\hat{\mu}^*_{L1} = \hat{\mu}^*_1 + \hat{\beta}^*_1(\mu_{x1} - \hat{\mu}_{x1}) \text{ และ } \hat{\mu}^*_{L2} \approx \hat{\mu}_2 + \hat{\beta}_2(\mu_{x1} - \hat{\mu}_{x2})$$

$$\hat{\beta}^*_1 = \frac{\sum_{i=1}^{n_{1rr}}(x_{1i} - \hat{\mu}_{x1rr})(y_{1i} - \hat{\mu}_{1rr})}{\sum_{i=1}^{n_{1rr}}(x_{1i} - \hat{\mu}_{x1rr})^2} \text{ และ } \hat{\beta}_2 = \frac{\sum_{i=1}^{n_2}(x_{2i} - \hat{\mu}_{x2})(y_{2i} - \hat{\mu}_2)}{\sum_{i=1}^{n_2}(x_{2i} - \hat{\mu}_{x2})^2}$$

2. ความแปรปรวนแบบไม่มีเงื่อนไขของตัวประมาณค่าเฉลี่ยประชากร

$$V\left(\hat{\mu}^*_{Lpst}\right) \simeq \frac{1}{N_T^2}\left[W_1^2\mu_1^2 + \left(W_2 - 1\right)^2\mu_2^2\right]V\left(\hat{N}_2\right) + \hat{W}_1^2\frac{(k_1 - 1)}{n_1}W_{1nr}S^2_{z1nr}$$

$$+\left[\left(1 - \frac{n_T}{\hat{N}_T}\right)\frac{1}{n_T}\sum_{h=1}^{2}\hat{W}_h S^2_{zh} + \frac{1}{n_T^2}\sum_{h=1}^{2}\left(1 - \hat{W}_h\right)S^2_{zh}\right] \tag{61}$$

เมื่อ

$$S^2_{zh} = \frac{1}{N_h - 1}\sum_{i=1}^{N_h}\left[(y_{hi} - \mu_h) - \beta_h(x_{hi} - \mu_{xh})\right]^2$$

$$\beta_h = \frac{\sum_{i=1}^{N_h}(x_{hi}-\mu_{xh})(y_{hi}-\mu_h)}{\sum_{i=1}^{N_h}(x_{hi}-\mu_{xh})^2}$$

$$S^2_{z1nr} = \frac{1}{N_{1nr}-1}\sum_{i=1}^{N_{1nr}}\left[(y_{1i}-\mu_h)-\beta_{1nr}(x_{1i}-\mu_{x1nr})\right]^2$$

$$\beta_{1nr} = \frac{\sum_{i=1}^{N_{1nr}}(x_{1i}-\mu_{x1nr})(y_{1i}-\mu_{1nr})}{\sum_{i=1}^{N_{1nr}}(x_{1i}-\mu_{x1nr})^2}$$

3. ค่าประมาณความแปรปรวนแบบไม่มีเงื่อนไขของตัวประมาณค่าเฉลี่ยประชากร

$$v\left(\hat{\mu}^*_{Lpst}\right) \simeq \frac{1}{\hat{N}_T^2}\left[\hat{W}_1^2\hat{\mu}^{*2}_{L1}+\left(\hat{W}_2-1\right)^2\hat{\mu}^{*2}_{L2}\right]v\left(\hat{N}_2\right)+\hat{W}_1^2\frac{(k_1-1)}{n_1}w_{1nr}s^2_{z1nr}$$

$$+\left[\left(1-\frac{n_T}{\hat{N}_T}\right)\frac{1}{n_T}\sum_{h=1}^{2}\hat{W}_h s^{*2}_{zh}+\frac{1}{n_T^2}\sum_{h=1}^{2}\left(1-\hat{W}_h\right)s^{*2}_{zh}\right] \qquad (62)$$

เมื่อ

$$\hat{\beta}_{1rr} = \frac{\sum_{i=1}^{n_{1rr}}(x_{1i}-\hat{\mu}_{x1rr})(y_{1i}-\hat{\mu}_{1rr})}{\sum_{i=1}^{n_{1rr}}(x_{1i}-\hat{\mu}_{x1rr})^2} \quad \text{และ} \quad \hat{\beta}'_{1nr} = \frac{\sum_{i=1}^{n'_{1nr}}(x_{1i}-\hat{\mu}'_{x1nr})(y_{1i}-\hat{\mu}'_{1nr})}{\sum_{i=1}^{n'_{1nr}}(x_{1i}-\hat{\mu}'_{x1nr})^2}$$

$$s^{*2}_{z1} = \frac{1}{n_{1rr}-2}\left[\sum_{i=1}^{n_{1rr}}(y_{1i}-\hat{\mu}_{1rr})^2-\hat{\beta}^2_{1rr}\sum_{i=1}^{n_{1rr}}(x_{1i}-\hat{\mu}_{x1rr})^2\right]$$

$$s^2_{z2} = \frac{1}{n_2-2}\left[\sum_{i=1}^{n_2}(y_{2i}-\hat{\mu}_2)^2-\hat{\beta}_2^{*2}\sum_{i=1}^{n_2}(x_{2i}-\hat{\mu}_{x2})^2\right]$$

$$s_{\varepsilon 1nr}^{2} = \frac{1}{n'_{1nr}-2}\left[\sum_{i=1}^{n'_{1nr}}(y_{1i}-\hat{\mu}'_{1nr})^{2}-\hat{\beta}'^{2}_{1nr}\sum_{i=1}^{n'_{1nr}}(x_{1i}-\hat{\mu}'_{x1nr})^{2}\right]$$

ตัวอย่าง 10 เพื่อให้เข้าใจง่ายสำหรับการสุ่มตัวอย่างเมื่อเกิดปัญหาการไม่ได้รับความร่วมมือ จากตัวอย่าง 1 โดยกรอบตัวอย่างไม่ครอบคลุม (ดังแสดงในภาพ 2 หน้า 44) มีขนาดประชากร N_1 = 70 มีขั้นตอนดังนี้

ขั้นตอน 1 สุ่มตัวอย่างแบบง่ายไม่แทนที่ขนาด n_1 จากประชากรในกรอบตัวอย่างไม่ครอบคลุม N_1 หน่วย และสำรวจหน่วยตัวอย่างที่ไม่อยู่ในกรอบตัวอย่างด้วย

จากภาพ 2 สมมติว่าสุ่มตัวอย่างแบบง่ายไม่แทนที่ขนาด 30% ของประชากรในกรอบตัวอย่างไม่ครอบคลุมขนาด N_1 = 70 ดังนั้น n_1 = 21 และสมมติว่าเกิดปัญหาการไม่ได้รับความร่วมมือ 30% ดังนั้นเราจะได้ตัวอย่างที่อยู่ในกรอบตัวอย่างให้ความร่วมมือ n_{1r} = 14 หน่วย และตัวอย่างที่อยู่ในกรอบตัวอย่างไม่ให้ความร่วมมือ n_{1nr} = 7 หน่วย และในระหว่างการสำรวจตัวอย่างขนาด n_1 = 21 พบว่า มีหน่วยตัวอย่างที่ไม่อยู่ในกรอบตัวอย่าง n_2 = 9 ดังนั้น $n_T = n_1 + n_2$ = 21 + 9 = 30 (ดังแสดงในตาราง 6) ซึ่งค่าสัมประสิทธิ์สหสัมพันธ์ของตัวอย่างที่ไม่อยู่ในกรอบตัวอย่างขนาด n_2 = 9 เท่ากับ 0.858

ตาราง 6 แสดงตัวอย่างข้อมูลที่ได้จากการสุ่มตัวอย่าง เมื่อเกิดปัญหากรอบตัวอย่างไม่ครอบคลุมและการไม่ได้รับความร่วมมือจากตัวอย่างที่อยู่ในกรอบตัวอย่าง

ตัวอย่างที่อยู่ในกรอบตัวอย่าง (h =1)				**ตัวอย่างที่ไม่อยู่ในกรอบตัวอย่าง (h =2)**		
หน่วยที่ (i)	m_{1i}	y_{1i}	x_{1i}	**หน่วยที่**	y_{2i}	x_{2i}
47	0	5.3	4	57_1	5.4	4
36	0	4.7	4	57_2	5.5	4
57	3	4.4	4	57_3	5.7	5
9	1	(4.0)	3	9_1	6.0	5
43	0	4.8	4	1_1	4.5	3
13	0	4.6	4	1_2	6.3	5
21	0	4.9	4	39_1	4.3	3
31	0	(3.8)	2	39_2	5.5	4
17	0	(4.4)	3	68_1	5.2	4
1	2	5.6	5			
5	0	5.1	5			

ตาราง 6 (ต่อ)

ตัวอย่างที่อยู่ในกรอบตัวอย่าง (h =1)				ตัวอย่างที่ไม่อยู่ในกรอบตัวอย่าง (h =2)		
หน่วยที่ (i)	m_{1i}	y_{1i}	x_{1i}	**หน่วยที่**	y_{2i}	x_{2i}
29	0	5.5	4			
55	0	6.8	6			
41	0	(6.1)	5			
27	0	5.0	4			
53	0	(6.8)	6			
39	2	(5.2)	4			
49	0	6.0	5			
3	0	3.3	2			
62	0	(3.8)	2			
68	1	4.2	4			
ผลรวม	9	y_{1r} =70.2	x_{1r} =59	**ผลรวม**	y_2 =48.4	x_2 =37
		(y_{1nr} =34.1)	x_{1nr} =25			

หมายเหตุ: ตัวเลขที่อยู่ในวงเล็บ แทน ข้อมูลที่ไม่ได้รับความร่วมมือในการให้ข้อมูล
หน่วยที่ 57_1 แทน ตัวอย่างที่ไม่อยู่ในกรอบตัวอย่างลำดับที่ 1 ที่อยู่ระหว่างหน่วยตัวอย่างที่อยู่ในกรอบตัวอย่างลำดับที่ i = 57 และ $i+1$ = 58

ขั้นตอน 2 สุ่มตัวอย่างย่อยในกลุ่มที่ไม่ความร่วมมือขนาด $n'_{1nr} = n_{1nr}/k_1$ = 7/2 ≈ 4 (กำหนด k_1 = 2) จากตัวอย่างที่ไม่ให้ความร่วมมือขนาด n_{1nr} = 7 (ดังแสดงในตาราง 7)

จากตัวอย่างที่ให้ความร่วมมือในการสุ่มตัวอย่างครั้งที่ 2 ขนาด n'_{1nr} = 4 มีค่าสัมประสิทธิ์สหสัมพันธ์ เท่ากับ 0.998 และค่าสัมประสิทธิ์สหสัมพันธ์ของตัวอย่างที่ให้ความร่วมมือทั้งหมดขนาด n_{1r} = 14 + 4 = 18 มีค่าเท่ากับ 0.914

ตาราง 7 แสดงตัวอย่างข้อมูลที่ได้จากการสุ่มตัวอย่างย่อยในกลุ่มตัวอย่างที่อยู่ในกรอบตัวอย่างไม่ครอบคลุมที่ไม่ให้ความร่วมมือ

หน่วยที่ (i)	y_{1i}	x_{1i}
17	4.4	3
31	3.8	2
41	6.1	5
53	6.8	6
ผลรวม	y'_{1nr} = 21.1	x'_{1nr} = 16

ตัวอย่าง 11 จากทฤษฎี 4 จะได้ตัวประมาณจำนวนประชากรที่ไม่อยู่ในกรอบตัวอย่าง มีค่าเท่ากับ

$$\hat{N}_2 = N_1 \frac{1}{n_1}\sum_{i=1}^{n_1} m_{1i} = (70)\frac{1}{21}(0+0+...+0+1) = (70)\frac{1}{21}(9) = 30$$

และจากทฤษฎี 5 จะได้ตัวประมาณจำนวนประชากร มีค่าเท่ากับ

$$\hat{N}_T = N_1 + \hat{N}_2 = 70 + 30 = 100$$

คำนวณค่าต่าง ๆ ที่เกี่ยวข้องได้ดังนี้

$$\bar{m} = \frac{1}{n_1}\sum_{i=1}^{n_1} m_{1i} = \frac{1}{21}(0+0+...+0+1) = \frac{1}{21}(9) = 0.4286$$

$$s_m^2 = \frac{1}{n_1 - 1}\sum_{i=1}^{n_1}(m_{1i} - \bar{m})^2 = \frac{1}{21\text{-}1}15.1429 = 0.7571$$

$$v\left(\hat{N}_2\right) = N_1^2\left(1 - \frac{n_1}{N_1}\right)\frac{s_m^2}{n_1} = (70)^2\left(1\text{-}\frac{21}{70}\right)\frac{0.7571}{21} = 123.6597$$

$$k_1 = \frac{n_{1nr}}{n'_{1nr}} = \frac{7}{4} = 1.75 \text{ และ } w_{1nr} = \frac{n_{1nr}}{n_1} = \frac{7}{21}$$

$$\hat{\mu}_1^* = \frac{1}{n_1}(y_{1r} + k_1 y'_{1nr}) = \frac{1}{21}[70.2+1.75(21.1)] = 5.1012$$

$$\hat{\mu}_2 = \frac{1}{n_2}\sum_{i=1}^{n_2} y_{2i} = \frac{1}{9}(48.4) = 5.3778$$

$$\hat{\mu}_{1rr} = \frac{1}{n_{1rr}}\left(y_{1r} + y'_{1nr}\right) = \frac{1}{14+4}(70.2+21.1) = 5.0722$$

$$s_1^{*2} = \frac{1}{n_{1rr}-1}\sum_{i=1}^{n_{1rr}}(y_i - \hat{\mu}_{1rr})^2 = \frac{1}{18\text{-}1}(15.2961) = 0.8998$$

$$\hat{\mu}'_{1nr} = \frac{y'_{1nr}}{n'_{1nr}} = \frac{21.1}{4} = 5.2750$$

$$s_{1nr}^2 = \frac{1}{n'_{1nr}-1}\sum_{i=1}^{n'_{1nr}}(y_{1i} - \hat{\mu}'_{1nr})^2 = \frac{1}{4\text{-}1}(5.9475) = 1.9825$$

$$s_2^2 = \frac{1}{n_2-1}\sum_{i=1}^{n_2}(y_{2i} - \hat{\mu}_2)^2 = \frac{1}{9\text{-}1}(3.3356) = 0.4170$$

จากทฤษฎี 12 จะได้ประมาณค่าเฉลี่ยที่เอนเอียงแบบไม่มีเงื่อนไขของค่าเฉลี่ยประชากร μ_T มีค่าเท่ากับ

$$\hat{\mu}_{pst}^* = \hat{W}_1\hat{\mu}_1^* + \hat{W}_2\hat{\mu}_2^* = \frac{70}{100}(5.1012) + \frac{30}{100}(5.3778) = 5.1842$$

มีค่าเอนเอียง เท่ากับ

$$Bias(\hat{\mu}_{pst}^*) = \hat{\mu}_{pst}^* - \mu_T = 5.1842 - 5.0130 = 0.1712$$

ค่าประมาณความแปรปรวนแบบไม่มีเงื่อนไขของตัวประมาณค่าเฉลี่ย มีค่าเท่ากับ

$$v\left(\hat{\mu}^{*}_{pst}\right) \simeq \frac{1}{\hat{N}_T^2}\left[\hat{W}_1^2\hat{\mu}_1^{*2} + \left(\hat{W}_2 - 1\right)^2 \hat{\mu}_2^{*2}\right] v\left(\hat{N}_2\right)$$

$$+\left(1-\frac{n_T}{\hat{N}_T}\right)\frac{1}{n_T}\sum_{h=1}^{2}\hat{W}_h s_h^{*2} + \frac{1}{n_T^2}\sum_{h=1}^{2}\left(1-\hat{W}_h\right)s_h^{*2} + \hat{W}_1^2\frac{\left(k_1-1\right)}{n_1}w_{1nr}s_{1nr}^2$$

$$\simeq \frac{1}{(100)^2}\left[\left(\frac{70}{100}\right)^2(5.1012)^2 + \left(\frac{30}{100}-1\right)^2(5.3778)^2\right](123.6597)+$$

$$\left(1-\frac{30}{100}\right)\frac{1}{30}\left[\left(\frac{70}{100}\right)0.8998+\left(\frac{30}{100}\right)0.4170\right]+$$

$$\frac{1}{(30)^2}\left[\left(1-\frac{70}{100}\right)0.8998+\left(1-\frac{30}{100}\right)0.4170\right]+$$

$$\left(\frac{70}{100}\right)^2\frac{1.75-1}{21}\left(\frac{7}{21}\right)(1.9825)$$

$$\simeq 0.3329 + 0.0176 + 0.0006 + 0.0116 = 0.3627$$

และค่าความคลาดเคลื่อนกำลังสองเฉลี่ย มีค่าเท่ากับ

$$MSE(\hat{\mu}^{*}_{pst}) = \left[Bias(\hat{\mu}^{*}_{pst})\right]^2 + v(\hat{\mu}^{*}_{pst}) = (0.1712)^2 + 0.3627 = 0.3920$$

ตัวอย่าง 12 จากทฤษฎี 13 จะได้ตัวประมาณค่าเฉลี่ยที่เอนเอียงแบบไม่มีเงื่อนไขของค่าเฉลี่ยประชากร μ_T มีค่าเท่ากับ

$$\hat{\mu}^*_{Rpst} = \hat{W}_1\hat{\mu}^*_{R1} + \hat{W}_2\hat{\mu}^*_{R2} = \frac{70}{100}(5.0101) + \frac{30}{100}(5.1390) = 5.0487$$

คำนวณค่าต่าง ๆ ได้ดังนี้

$$\hat{\mu}_{x1} = \frac{84}{21} = 4.00,\ \hat{\mu}_{x2} = \frac{37}{9} = 4.1111$$

$$\hat{R}^*_1 = \frac{\hat{\mu}^*_1}{\hat{\mu}_{x1}} = \frac{5.1012}{4.0000} = 1.2753 \text{ และ } \hat{\mu}^*_{R1} = \hat{R}^*_1\mu_{x1} = 1.2753(3.9286) = 5.0101$$

$$\hat{R}_2 = \frac{\hat{\mu}_2}{\hat{\mu}_{x2}} = \frac{48.4/9}{37/9} = 1.3081 \text{ และ } \hat{\mu}^*_{R2} \approx \hat{R}_2\mu_{x1} = 1.3081(3.9286) = 5.1390$$

$$\hat{\mu}_{x1rr} = \frac{1}{n_{1rr}}\left(x_{1r} + x'_{1nr}\right) = \frac{1}{14+4}(59+16) = \frac{75}{18} = 4.1667$$

$$\hat{R}_{1rr} = \frac{\hat{\mu}_{1rr}}{\hat{\mu}_{x1rr}} = \frac{91.3/18}{75/18} = 1.2173$$

$$s^{*2}_{d1} = \frac{1}{n_{1rr}-1}\sum_{i=1}^{n_{1rr}}(y_{1i} - \hat{R}_{1rr}x_{1i})^2 = \frac{1}{18-1}(6.2740) = 0.3691$$

$$\hat{\mu}'_{1nr} = \frac{y'_{1nr}}{n'_{1nr}} = \frac{21.1}{4} = 5.2750 \text{ และ } \hat{\mu}'_{x1nr} = \frac{x'_{1nr}}{n'_{1nr}} = \frac{16}{4} = 4.0$$

$$\hat{R}'_{1nr} = \frac{\hat{\mu}'_{1nr}}{\hat{\mu}'_{x1nr}} = \frac{21.1/4}{16/4} = 1.3188$$

$$s^2_{d1nr} = \frac{1}{n'_{1nr}-1}\sum_{i=1}^{n'_{1nr}}(y_i - \hat{R}'_{1nr}x_{1i})^2 = \frac{1}{4-1}(3.0298) = 1.0099$$

$$\hat{R}_2 = \frac{\hat{\mu}_2}{\hat{\mu}_{x2}} = \frac{48.4/9}{39/9} = 1.2410$$

$$s_{d2}^{*2} = s_{d2}^2 = \frac{1}{n_2 - 1}\sum_{i=1}^{n_2}(y_i - \hat{R}_2 x_{2i})^2 = \frac{1}{9\text{-}1}(1.7014) = 0.2127$$

มีค่าเอนเอียง เท่ากับ

$$Bias(\hat{\mu}_{Rpst}^*) = \hat{\mu}_{Rpst}^* - \mu_T = 5.0487 - 5.0130 = 0.0357$$

ค่าประมาณความแปรปรวนแบบไม่มีเงื่อนไขของตัวประมาณค่าเฉลี่ย มีค่าเท่ากับ

$$v\left(\hat{\mu}_{Rpst}^*\right) \simeq \frac{1}{\hat{N}_T^2}\left[\hat{W}_1^2\hat{\mu}_{R1}^{*2} + \left(\hat{W}_2 - 1\right)^2\hat{\mu}_{R2}^{*2}\right]v\left(\hat{N}_2\right)$$

$$+\left(1 - \frac{n_T}{\hat{N}_T}\right)\frac{1}{n_T}\sum_{h=1}^{2}\hat{W}_h s_{dh}^{*2} + \frac{1}{n_T^2}\sum_{h=1}^{2}\left(1 - \hat{W}_h\right)s_{dh}^{*2} + \hat{W}_1^2\frac{\left(k_1 - 1\right)}{n_1}w_{1nr}s_{d1nr}^2$$

$$\simeq \frac{1}{(100)^2}\left[\left(\frac{70}{100}\right)^2(5.0101)^2 + \left(\frac{30}{100}\text{-}1\right)^2(5.1390)^2\right](123.6597) +$$

$$\left(1\text{-}\frac{30}{100}\right)\frac{1}{30}\left[\left(\frac{70}{100}\right)0.3691 + \left(\frac{30}{100}\right)0.2127\right] +$$

$$\frac{1}{(30)^2}\left[\left(1\text{-}\frac{70}{100}\right)0.3691 + \left(1\text{-}\frac{30}{100}\right)0.2127\right] +$$

$$\left(\frac{70}{100}\right)^2\frac{1.75\text{-}1}{21}\left(\frac{7}{21}\right)(1.0099)$$

$$\simeq 0.3121 + 0.0075 + 0.0003 + 0.0116 = 0.3315$$

และค่าความคลาดเคลื่อนกำลังสองเฉลี่ย มีค่าเท่ากับ

$$MSE(\hat{\mu}^{*}_{Rpst}) = \left[Bias(\hat{\mu}^{*}_{Rpst}) \right]^2 + v(\hat{\mu}^{*}_{Rpst}) = (0.0357)^2 + 0.3315 = 0.3328$$

ตัวอย่าง 13 จากทฤษฎี 14 จะได้ตัวประมาณค่าเฉลี่ยที่เอนเอียงแบบไม่มีเงื่อนไขของค่าเฉลี่ยประชากร μ_T มีค่าเท่ากับ

$$\hat{\mu}^{*}_{Lpst} = \hat{W}_1 \hat{\mu}^{*}_{L1} + \hat{W}_2 \hat{\mu}^{*}_{L2} = \frac{70}{100}(5.0448) + \frac{30}{100}(5.3725) = 5.1431$$

คำนวณค่าต่าง ๆ ได้ดังนี้

$$\hat{\beta}^{*}_1 \approx \hat{\beta}_{1rr} = \frac{\sum_{i=1}^{n_{1rr}} (x_{1i} - \hat{\mu}_{x1rr})(y_{1i} - \hat{\mu}_{1rr})}{\sum_{i=1}^{n_{1rr}} (x_{1i} - \hat{\mu}_{x1rr})^2} = \frac{16.1833}{20.5} = 0.7894$$

$$\hat{\beta}'_{1nr} = \frac{\sum_{i=1}^{n'_{1nr}} (x_{1i} - \hat{\mu}'_{x1nr})(y_{1i} - \hat{\mu}'_{1nr})}{\sum_{i=1}^{n'_{1nr}} (x_{1i} - \hat{\mu}'_{x1nr})^2} = \frac{7.7}{10} = 0.77$$

$$\hat{\beta}^{*}_2 = \frac{\sum_{i=1}^{n_2} (x_{2i} - \hat{\mu}_{x2})(y_{2i} - \hat{\mu}_2)}{\sum_{i=1}^{n_2} (x_{2i} - \hat{\mu}_{x2})^2} = \frac{3.8222}{130.8105} = 0.0292$$

$$\hat{\mu}^{*}_{L1} = \hat{\mu}^{*}_1 + \hat{\beta}^{*}_1(\mu_{x1} - \hat{\mu}_{x1}) = 5.1012 + 0.7894(3.9286 - 4.00) = 5.0448$$

$$\hat{\mu}^{*}_{L2} = \hat{\mu}^{*}_2 + \hat{\beta}^{*}_2(\mu_{x2} - \hat{\mu}_{x2}) \approx \hat{\mu}^{*}_2 + \hat{\beta}^{*}_2(\mu_{x1} - \hat{\mu}_{x2})$$

$$= 5.3778 + 0.0292\,(3.9286 - 4.1111) = 5.3725$$

$$s_{z1}^2 = \frac{1}{n_{1rr}-2}\left[\sum_{i=1}^{n_{1rr}}(y_{1i}-\hat{\mu}_{1rr})^2 - \hat{\beta}_{1rr}^2\sum_{i=1}^{n_{1rr}}(x_{1i}-\hat{\mu}_{x1rr})^2\right] = \frac{1}{18\text{-}2}(2.5205) = 0.1575$$

$$s_{z2}^2 = \frac{1}{n_2-2}\left[\sum_{i=1}^{n_2}(y_{2i}-\hat{\mu}_2)^2 - \hat{\beta}_2^{*2}\sum_{i=1}^{n_2}(x_{2i}-\hat{\mu}_{x2})^2\right] = \frac{1}{9\text{-}2}(3.2239) = 0.2015$$

$$s_{z1nr}^2 = \frac{1}{n'_{1nr}-2}\left[\sum_{i=1}^{n'_{1nr}}(y_{1i}-\hat{\mu}'_{1nr})^2 - \hat{\beta}'^2_{1nr}\sum_{i=1}^{n'_{1nr}}(x_{1i}-\hat{\mu}'_{x1nr})^2\right] = \frac{1}{4\text{-}2}(0.0185) = 0.0093$$

มีค่าเอนเอียง เท่ากับ

$$Bias(\hat{\mu}^*_{Lpst}) = \hat{\mu}^*_{Lpst} - \mu_T = 5.1431 - 5.0130 = 0.1301$$

ค่าประมาณความแปรปรวนแบบไม่มีเงื่อนไขของตัวประมาณค่าเฉลี่ย มีค่าเท่ากับ

$$v\left(\hat{\mu}^*_{Lpst}\right) \simeq \frac{1}{\hat{N}_T^2}\left[\hat{W}_1^2\hat{\mu}_{L1}^{*2} + \left(\hat{W}_2-1\right)^2\hat{\mu}_{L2}^{*2}\right]v\left(\hat{N}_2\right)$$

$$+\left[\left(1-\frac{n_T}{\hat{N}_T}\right)\frac{1}{n_T}\sum_{h=1}^{2}\hat{W}_h s_{zh}^{*2} + \frac{1}{n_T^2}\sum_{h=1}^{2}\left(1-\hat{W}_h\right)s_{zh}^{*2}\right] + \hat{W}_1^2\frac{(k_1-1)}{n_1}w_{1nr}s_{z1nr}^2$$

$$\simeq \frac{1}{(100)^2}\left[\left(\frac{70}{100}\right)^2(5.0448)^2 + \left(\frac{30}{100}\text{-}1\right)^2(5.3725)^2\right](123.6597) +$$

$$\left(1\text{-}\frac{30}{100}\right)\frac{1}{30}\left[\left(\frac{70}{100}\right)0.1575 + \left(\frac{30}{100}\right)0.2015\right] +$$

$$\frac{1}{(30)^2}\left[\left(1\text{-}\frac{70}{100}\right)0.1575 + \left(1\text{-}\frac{30}{100}\right)0.2015\right] +$$

$$\left(\frac{70}{100}\right)^2 \frac{1.75-1}{21}\left(\frac{7}{21}\right)(0.0093)$$

$$\simeq 0.3291 + 0.0040 + 0.0002 + 0.0116 = 0.3449$$

และค่าความคลาดเคลื่อนกำลังสองเฉลี่ย มีค่าเท่ากับ

$$MSE(\hat{\mu}^*_{Lpst}) = \left[Bias(\hat{\mu}^*_{Lpst})\right]^2 + v(\hat{\mu}^*_{Lpst}) = (0.1301)^2 + 0.3618 = 0.3618$$

จากตัวอย่าง 11-13 จะได้

$$MSE(\hat{\mu}^*_{Rpst}) = 0.3328 < MSE(\hat{\mu}^*_{Lpst}) = 0.3618 < MSE(\hat{\mu}^*_{pst}) = 0.3920$$

การเปรียบเทียบประสิทธิภาพของตัวประมาณ

ในการประมาณค่าพารามิเตอร์ เช่น ค่าเฉลี่ย อาจใช้ตัวประมาณค่าต่างกันได้ จึงน่าจะที่พิจารณาว่าควรจะใช้ตัวประมาณใดถึงจะดีที่สุด การเลือกใช้ตัวประมาณควรมีหลักเกณฑ์ในการเลือกอย่างเหมาะสม หากเป็นบุคคลทั่วไปอาจต้องการใช้ตัวประมาณที่คิดได้ง่าย ๆ แต่สำหรับทางวิชาการมีหลักเกณฑ์ที่ใช้อยู่หลายประการ เช่น ความไม่เอนเอียง ความคงเส้นคงวา และความคลาดเคลื่อนต่ำ เป็นต้น สมบัติที่สำคัญ 2 ประการของตัวประมาณที่มักใช้ในการสำรวจด้วยตัวอย่าง ได้แก่ ความไม่เอนเอียง และการมีความคลาดเคลื่อนต่ำ (ประชุม สุวัตถี, 2552, หน้า 69-72) ในการเปรียบเทียบตัวประมาณ $\hat{\theta}_1$ และ $\hat{\theta}_2$ ของพารามิเตอร์ θ จึงควรพิจารณาคุณสมบัติของ $\hat{\theta}_1$ และ $\hat{\theta}_2$ ดังนี้

1. **ความเอนเอียง** (Bias) โดยทั่วไปเราต้องการให้ตัวประมาณมีความเอนเอียงต่ำ หรือเมื่อเปรียบเทียบตัวประมาณ $\hat{\theta}_1$ กับ $\hat{\theta}_2$ ของ θ ที่มาจากตัวอย่างเดียวกัน ซึ่งตัวประมาณที่มีความเอนเอียง หรืออาจพิจารณาความเอนเอียงสัมพัทธ์ (Relative Bias) ต่ำกว่าอีกตัวหนึ่ง เป็นตัวประมาณที่ดีกว่าอีกตัวหนึ่ง

บางครั้งการพิจารณาความเอนเอียงของ $\hat{\theta}_1$ และ $\hat{\theta}_2$ ทำได้ยาก จึงอาจใช้การจำลอง (Simulation) มาช่วย และคำนวณความเอนเอียง และความเอนเอียงสัมพัทธ์ ดังนี้

ความเอนเอียง:

$$Bias(\hat{\theta}_1) = \hat{\theta}_1 - \theta \text{ และ } Bias(\hat{\theta}_2) = \hat{\theta}_2 - \theta$$

ความเอนเอียงสัมพัทธ์:

$$RB(\hat{\theta}_1) = \frac{Bias(\hat{\theta}_1)}{\theta} \text{ และ } RB(\hat{\theta}_2) = \frac{Bias(\hat{\theta}_2)}{\theta}$$

2. **ความแม่นยำ** (Accuracy) ในกรณีที่ตัวประมาณ $\hat{\theta}_1$ และ $\hat{\theta}_2$ ของพารามิเตอร์ θ เป็นตัวประมาณเอนเอียงในการเปรียบเทียบ $\hat{\theta}_1$ กับ $\hat{\theta}_2$ นอกจากจะพิจารณาเปรียบเทียบความเอนเอียงหรือความเอนเอียงสัมพัทธ์แล้ว ผู้ศึกษาอาจพิจารณาความแม่นยำ หรือความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) ของตัวประมาณ หรือพิจารณาความแม่นยำสัมพัทธ์ (Relative Accuracy) ซึ่งเราต้องการตัวประมาณที่มีความแม่นยำหรือความแม่นยำสัมพัทธ์ต่ำ ซึ่งแปลว่ามีความคลาดเคลื่อนต่ำ โดยที่ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) และความแม่นยำสัมพัทธ์คำนวณได้ดังนี้

ความแม่นยำหรือความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE):

$$MSE(\hat{\theta}_1) = \left[Bias(\hat{\theta}_1)\right]^2 + V(\hat{\theta}_1) \text{ และ } MSE(\hat{\theta}_2) = \left[Bias(\hat{\theta}_2)\right]^2 + V(\hat{\theta}_2)$$

ความแม่นยำสัมพัทธ์:

$$RA(\hat{\theta}_1) = \frac{MSE(\hat{\theta}_1)}{\theta} \text{ และ } RA(\hat{\theta}_2) = \frac{MSE(\hat{\theta}_2)}{\theta}$$

ซึ่ง $\hat{\theta}_1$ จะมีประสิทธิภาพดีกว่า $\hat{\theta}_2$ ด้านความแม่นยำถ้า $MSE(\hat{\theta}_1) \le MSE(\hat{\theta}_2)$ หรือ $\hat{\theta}_1$ จะมีประสิทธิภาพดีกว่า $\hat{\theta}_2$ ด้านความแม่นยำสัมพัทธ์ ถ้า $RA(\hat{\theta}_1) \le RA(\hat{\theta}_2)$

3. **ความเที่ยงตรง** (Precision) กรณีที่ตัวประมาณไม่เอนเอียงจะใช้ความแปรปรวนของตัวประมาณเป็นเครื่องมือวัดความเที่ยงตรงของตัวประมาณนั้น และถือว่าตัวประมาณที่มีความแปรปรวนต่ำที่สุดเป็นตัวประมาณที่ดีที่สุด (Best Estimator) บางครั้งถึงแม้ว่าตัวประมาณเป็นตัว

ประมาณเอนเอียง แต่ยังใช้ความแปรปรวนเป็นเครื่องมือวัดความเที่ยงตรงอยู่ดี ซึ่งความใช้ได้ถ้าความเอนเอียงไม่มากเกินไป มักเกิดขึ้นเมื่อตัวอย่างมีขนาดใหญ่

เมื่อต้องการเปรียบเทียบตัวประมาณ $\hat{\theta}_1$ และ $\hat{\theta}_2$ ของพารามิเตอร์ θ เมื่อทั้งสองตัวเป็นตัวประมาณไม่เอนเอียงหรือเอนเอียงน้อย เราจะถือว่าตัวประมาณที่มีความแปรปรวนต่ำกว่าเป็นตัวประมาณที่ดีกว่าอีกตัวหนึ่ง บางครั้งอาจพิจารณาประสิทธิภาพสัมพัทธ์ (Relative Efficiency) ของ $\hat{\theta}_1$ เทียบกับ $\hat{\theta}_2$ ซึ่งได้แก่

$$RE(\hat{\theta}_1 / \hat{\theta}_2) = \frac{V(\hat{\theta}_2)}{V(\hat{\theta}_1)}$$

ถ้า $RE(\hat{\theta}_1 / \hat{\theta}_2) > 1$ แปลว่า $\hat{\theta}_1$ ดีกว่า $\hat{\theta}_2$

เมื่อใช้ข้อมูลจากตัวอย่าง อาจประมาณประสิทธิภาพของ $\hat{\theta}_1$ เทียบกับ $\hat{\theta}_2$ ได้ด้วย

$$\widehat{RE}(\hat{\theta}_1 / \hat{\theta}_2) = \frac{v(\hat{\theta}_2)}{v(\hat{\theta}_1)}$$

การวิจัยครั้งนี้ ใช้เกณฑ์ค่าความคลาดเคลื่อนกำลังสองต่ำสุด เป็นเกณฑ์ในการเปรียบเทียบประสิทธิภาพของตัวประมาณค่าเฉลี่ยประชากร

งานวิจัยที่เกี่ยวข้อง

Okafor and Lee (2000) เสนอการประมาณค่าเฉลี่ยด้วยตัวประมาณอัตราส่วน และตัวประมาณการถดถอยในแผนการสุ่มตัวอย่างแบบสองขั้นตอน เมื่อมีการสุ่มตัวอย่างย่อยในกลุ่มที่ไม่ให้ความร่วมมือ โดยมีเงื่อนไขว่าไม่ทราบตัวแปรช่วยที่คาดว่ามีความสัมพันธ์กับตัวแปรที่ศึกษา จึงสุ่มตัวอย่าง 2 ครั้ง โดยครั้งแรกทำการสุ่มตัวอย่างเพื่อศึกษาตัวแปรช่วย จากนั้นจึงสุ่มตัวอย่างครั้งที่สอง ซึ่งเป็นการสุ่มซ้ำบางส่วนของตัวอย่างในครั้งแรกเพื่อศึกษาตัวแปรที่ศึกษา และในการสุ่มตัวอย่างครั้งที่สองนี้เกิดปัญหาการไม่ได้รับความร่วมมือจากบางหน่วยของตัวอย่าง จึงทำการสุ่มตัวอย่างย่อยกลุ่มตัวอย่างที่ไม่ให้ความร่วมมืออีกครั้ง

Siripornpibul (2001) ได้เสนอการประมาณค่าเฉลี่ยด้วยตัวประมาณอย่างง่าย เมื่อมีการสุ่มตัวอย่างย่อยในกลุ่มที่ไม่ให้ความร่วมมือ 2 ครั้ง ซึ่งเสนอแนวทางแก้ไขปัญหาการไม่ให้ความ

ร่วมมือ โดยใช้สาระโดยตรงจากกลุ่มที่ไม่ให้ความร่วมมือ ด้วยการสุ่มย่อยในกลุ่มที่ไม่ให้ความร่วมมือ เพื่อเปลี่ยนผู้ที่ไม่ให้ความร่วมมือให้กลายเป็นผู้ให้ความร่วมมือ

Ängsved (2004) ได้เสนอการประมาณค่าผลรวมประชากรด้วยการประมาณค่าเฉพาะส่วนหลัก (Estimation of Domain) นั่นคือ ใช้เฉพาะข้อมูลของตัวอย่างที่ได้จากกรอบตัวอย่างในการประมาณค่าผลรวมประชากร เมื่อประชากรที่อยู่ในกรอบตัวอย่างและประชากรที่ไม่อยู่ในกรอบตัวอย่างมีคุณลักษณะของค่าเฉลี่ยสอดคล้องกัน ซึ่งวิธีการนี้จะไม่เหมาะสมเมื่อประชากรที่อยู่ในกรอบตัวอย่างและประชากรที่ไม่อยู่ในกรอบตัวอย่างมีคุณลักษณะของค่าเฉลี่ยต่างกัน เนื่องจากจะส่งผลต่อการประมวลผลและวิเคราะห์ข้อมูล หรือค่าประมาณที่ได้จะมีค่าเอนเอียงไปจากค่าที่แท้จริงของประชากร และจะประมาณค่าเฉลี่ยประชากรที่ไม่อยู่ในกรอบตัวอย่าง μ_{x2} ด้วยค่าเฉลี่ยประชากรที่อยู่ในกรอบตัวอย่าง μ_{x1}

Agarwal and Gupta (2008) ได้เสนอการสุ่มตัวอย่างแบบง่ายไม่แทนที่เมื่อเกิดปัญหากรอบตัวอย่างไม่ครอบคลุม โดยมีการเก็บรวบรวมจำนวนหน่วยตัวอย่างที่ไม่อยู่ในกรอบตัวอย่างในระหว่างที่มีการสำรวจตัวอย่างจากกรอบตัวอย่าง และศึกษาตัวประมาณค่า N_2 ตลอดจนศึกษาการประมาณค่าเฉลี่ยประชากรด้วยตัวประมาณอย่างง่าย ซึ่งแยกพิจารณา 2 กรณี คือ

1. กรณีประชากรที่อยู่ในกรอบตัวอย่าง และประชากรที่ไม่อยู่ในกรอบตัวอย่างมีคุณลักษณะของค่าเฉลี่ยสอดคล้องกัน จะประมาณค่าเฉลี่ยประชากรด้วยค่าเฉลี่ยตัวอย่างที่ได้จากกรอบตัวอย่างร่วมกับการใช้ตัวประมาณจำนวนประชากรที่ไม่อยู่ในกรอบตัวอย่าง

2. กรณีประชากรที่อยู่ในกรอบตัวอย่าง และประชากรที่ไม่อยู่ในกรอบตัวอย่างมีคุณลักษณะของค่าเฉลี่ยต่างกัน จะประมาณค่าเฉลี่ยประชากรด้วยค่าเฉลี่ยตัวอย่างที่ได้จากกรอบตัวอย่าง และค่าเฉลี่ยตัวอย่างย่อยที่สุ่มจากกลุ่มตัวอย่างที่ไม่อยู่ในกรอบตัวอย่าง แต่จะพิจารณากรณีที่ทราบค่า N_2

Andy, et al. (2008) ได้ศึกษาความคลาดเคลื่อนจากการสำรวจทางโทรศัพท์ เนื่องจากปัญหากรอบตัวอย่างไม่สมบูรณ์ และการไม่ได้รับความร่วมมือพบว่าก่อนการสำรวจจะไม่ทราบว่าเกิดปัญหากรอบตัวอย่างไม่สมบูรณ์หรือปัญหาการไม่ได้รับความร่วมมือ จนกระทั่งภายหลังการสัมภาษณ์ทางโทรศัพท์ จึงจะสามารถจำแนกได้ว่าหน่วยตัวอย่างใดอยู่หรือไม่อยู่ในกรอบตัวอย่างและหน่วยตัวอย่างใดให้ความร่วมมือหรือไม่ให้ความร่วมมือ โดยใช้เทคนิคการแบ่งชั้นภูมิภายหลังการแจงนับ ผลการศึกษาพบว่า การสำรวจทางโทรศัพท์มักจะเจอปัญหาการไม่ได้รับความร่วมมือมากกว่าปัญหากรอบตัวอย่างไม่สมบูรณ์