บทที่ 2

เอกสาร และงานวิจัยที่เกี่ยวข้อง

ในการวิจัยครั้งนี้จะกล่าวถึงเอกสาร และงานวิจัยที่เกี่ยวข้องกับการเปรียบเทียบประสิทธิภาพวิธีประมาณค่าเฉลี่ยที่ดัดแปลงมาจากตัวประมาณ Horvitz-Thompson ภายใต้แผนการสุ่มตัวอย่างแบบกลุ่มปรับ เมื่อสุ่มหน่วยตัวอย่างอย่างง่าย โดยมีรายละเอียดดังนี้

1. ทฤษฎีการจำลองแบบประชากรที่มีลักษณะหายาก

   1.1 กระบวนการปัวซง

   1.2 กระบวนการปัวซงคลัสเตอร์

2. ทฤษฎีการสุ่มตัวอย่างแบบพื้นฐาน

   2.1 แผนการสุ่มตัวอย่างด้วยความน่าจะเป็นเท่ากัน

      2.1.1 การสุ่มตัวอย่างแบบง่าย

      2.1.2 การสุ่มตัวอย่างแบบมีชั้นภูมิ

      2.1.3 การสุ่มตัวอย่างแบบมีระบบ

      2.1.4 การสุ่มตัวอย่างแบบกลุ่ม

   2.2 แผนการสุ่มตัวอย่างแบบกลุ่มด้วยความน่าจะเป็นไม่เท่ากัน

   2.3 วิธีการประมาณค่าเฉลี่ยด้วยความน่าจะเป็นเท่ากัน

      2.3.1 การประมาณค่าเฉลี่ยแบบง่าย

      2.3.2 การประมาณค่าเฉลี่ยแบบอัตราส่วน

   2.4 วิธีการประมาณค่าเฉลี่ยด้วยความน่าจะเป็นไม่เท่ากัน

      2.4.1 การประมาณค่าเฉลี่ยอย่างง่าย

      2.4.2 การประมาณค่าเฉลี่ยแบบอัตราส่วน

3. ทฤษฎีการสุ่มตัวอย่างแบบกลุ่มปรับเมื่อสุ่มหน่วยตัวอย่างอย่างง่าย

   3.1 ตัวประมาณค่าเฉลี่ยแบบง่ายที่ดัดแปลงมาจากตัวประมาณ Horvitz–Thompson

   3.2 ตัวประมาณค่าเฉลี่ยแบบอัตราส่วนที่ดัดแปลงมาจากตัวประมาณ Horvitz–Thompson

   3.3 ตัวประมาณค่าเฉลี่ยแบบอัตราส่วนที่ปรับปรุงโดยวิธีการ Rao-Blackwell

3.4 ตัวอย่างการคำนวณ

4. งานวิจัยที่เกี่ยวข้อง

ทฤษฎีการจำลองแบบประชากรที่มีลักษณะหายาก

การจำลองแบบประชากรในงานวิจัยนี้เป็นการจำลองประชากรที่มีลักษณะหายาก และอยู่รวมกันเป็นกลุ่มซึ่งสามารถจำลองได้ด้วยกระบวนการปัวซงคลัสเตอร์ เพื่อนำประชากรที่จำลองได้ไปใช้ศึกษาเกี่ยวกับความแปรปรวนของตัวประมาณค่าเฉลี่ยของแผนการสุ่มตัวอย่างแบบกลุ่มปรับที่หน่วยตัวอย่างสุ่มด้วยวิธีการอย่างง่าย โดยทฤษฎีในส่วนนี้ประกอบด้วย กระบวนการปัวซง และ กระบวนการปัวซงคลัสเตอร์

1. **กระบวนการปัวซง**

กระบวนการปัวซงเป็นกระบวนการที่เป็นรูปแบบมาตรฐานสำหรับกระบวนการเชิงพื้นที่ ซึ่งเป็นกระบวนการที่เกิดขึ้นในชีวิตจริง ที่เกี่ยวกับระยะห่างของค่าสังเกตแต่ละตัว แต่ไม่มีรูปแบบที่เป็นมาตรฐานจึงมีการพัฒนาเพื่อต้องการให้มีการแจกแจงความถี่เป็นไปตามทฤษฎี และสอดคล้องกับสมมติฐานของกระบวนการ ดังนั้นรูปแบบที่เป็นมาตรฐานสำหรับกระบวนการเชิงพื้นที่ที่ใช้อธิบายการแจกแจงเชิงสุ่มของจุดต่าง ๆ ในพื้นที่หนึ่ง (A.D.Cliff and J.K. Ord, 1981, pp. 87-89) มีวิธีดำเนินการดังนี้

1.1 กำหนดกรอบพื้นที่ที่ได้รับการแบ่งเป็นพื้นที่ย่อยเล็ก ๆ N หน่วย และแต่ละพื้นที่ย่อยมีขนาด 1/N

1.2 สำหรับแต่ละพื้นที่ย่อยเล็ก ๆ กำหนดความน่าจะเป็นของการเกิดเหตุการณ์ที่เราสนใจดังนี้

p (no individuals in subarea) $= p_0 = 1 - \lambda / N + O(N^{-2})$

p (1 individuals in subarea) $= p_1 = \lambda / N + O(N^{-2})$

p (2 or more individuals in subarea) $= O(N^{-2})$

เมื่อ $O(N^{-2})$ หมายถึง ความน่าจะเป็นที่จะพบสิ่งที่สนใจมากกว่า 1 สิ่งในพื้นที่เล็ก ๆ ที่มีค่าเข้าใกล้ศูนย์

1.3 จำนวนสิ่งที่เราสนใจ X ในกรอบพื้นที่ที่ได้จากข้อ 1.2 เป็นไปตามการแจกแจงแบบ ทวินาม (Binomial Distribution) คือ

$$p(X = x) = \binom{N}{x} p_1^x p_0^{N-x}, \qquad x = 0,1,\ldots,N$$

1.4 ถ้าจำนวนพื้นที่ย่อยมากพอหรือ $N \to \infty$ โดยที่ $Np_1 = \lambda$ ซึ่งเป็นค่าคงที่ที่เป็นขีดจำกัดการแจกแจงความน่าจะเป็นของ X อยู่ในรูป

$$p(X = x) = \frac{\exp(-\lambda)\lambda^x}{x!}, \qquad x = 0,1,\ldots,$$

ซึ่งเป็นการแจกแจงแบบปัวซง โดยมีข้อสมมติที่สำคัญของกระบวนการปัวซงคือ

1.4.1 ไม่มีปฏิสัมพันธ์กันระหว่างพื้นที่ย่อยต่าง ๆ

1.4.2 ไม่มีความเป็นไปได้ของการรวมกลุ่มที่ซับซ้อนของสิ่งใด ๆ ภายในแต่ละพื้นที่ย่อย

1.4.3 ไม่มีแนวโน้มของพื้นที่ใกล้เคียงที่แสดงลักษณะคล้ายกัน

ในการสร้างกระบวนการปัวซงที่เกี่ยวกับพื้นที่หรือเวลาโดยมีพารามิเตอร์ λ จะมีผลลัพธ์เหมือนกัน ที่สำคัญ 2 ประการคือ

1. ระยะห่างของพื้นที่ที่พบสิ่งที่สนใจหรือของเวลาระหว่างเหตุการณ์ เป็นตัวแปรสุ่มที่เป็นอิสระกันจากฟังก์ชันการแจกแจงความน่าจะเป็น (p.d.f) แบบเอกซ์โปเนนเชียลที่มีพารามิเตอร์ λ

2. จำนวนครั้งของเหตุการณ์ ในพื้นที่ที่มีขนาด A หรือ ในช่วงเวลาที่กำหนด t มีการแจกแจงแบบพัวซองที่มีพารามิเตอร์ λA หรือ λt ตามลำดับ

2. กระบวนการปัวซงคลัสเตอร์

กระบวนการปัวซงคลัสเตอร์ หรือเรียกอีกอย่างหนึ่งว่า Centre-Satellite Process ได้รับการพัฒนาครั้งแรกโดย Neyman และ Scott (1958) ซึ่งต่อมา A.D.Cliff และ J.K.Ord (1981) ได้พัฒนา และสรุปกระบวนการปังซงคลัสเตอร์ซึ่งสร้างมาจากขั้นตอนที่สำคัญ 3 ขั้นตอน ดังนี้

2.1 จำนวนจุดหลักสร้างด้วยกระบวนการปัวซง ที่มีความหนาแน่น λ ต่อหนึ่งหน่วยพื้นที่

2.2 จำนวนบริวารของแต่ละจุดหลัก ซึ่งเป็นค่าสังเกตที่เป็นอิสระต่อกัน มาจากการแจกแจงความน่าจะเป็นที่กำหนด

2.3 ที่ตั้งของบริวารแต่ละตัวมีความสัมพันธ์กับตำแหน่งของจุดหลักซึ่งเป็นอิสระกัน โดยรายละเอียดได้แสดงในบทที่ 3

ทฤษฎีการสุ่มตัวอย่างแบบพื้นฐาน

ก่อนจะศึกษาเกี่ยวกับแผนการสุ่มตัวอย่างแบบกลุ่มปรับ ควรทำความเข้าใจเกี่ยวกับทฤษฎีการสุ่มตัวอย่างแบบพื้นฐาน ซึ่งได้แก่ แผนการสุ่มตัวอย่างโดยใช้ความน่าจะเป็น และวิธีการประมาณค่า ซึ่งมีรายละเอียดดังนี้

1. แผนการสุ่มตัวอย่างด้วยความน่าจะเป็นเท่ากัน

แผนการสุ่มตัวอย่างด้วยความน่าจะเป็นเท่ากัน คือการเลือกตัวอย่างจากประชากรหนึ่ง ที่ตัวอย่างที่เป็นไปได้แต่ละตัวอย่างมีความน่าจะเป็นที่จะถูกเลือกเท่ากัน (ประชุม สุวัตถี, 2552, หน้า 7) ซึ่งวิธีการสุ่มตัวอย่างมีหลายวิธี ที่เป็นหลักสำคัญได้แก่ การสุ่มตัวอย่างแบบง่าย การสุ่มตัวอย่างแบบมีชั้นภูมิ การสุ่มตัวอย่างแบบมีระบบ และการสุ่มตัวอย่างแบบกลุ่ม

1.1 การสุ่มตัวอย่างแบบง่าย

แผนการสุ่มตัวอย่างแบบง่ายเป็นวิธีการสุ่มตัวอย่างที่ทำให้แต่ละหน่วยตัวอย่างของประชากรมีความน่าจะเป็นในการถูกเลือกเท่ากัน และเหมาะกับประชากรที่มีลักษณะคล้ายกัน หรือกล่าวอีกนัยว่าควรเลือกใช้การสุ่มตัวอย่างแบบง่ายกับประชากรที่มีค่าความแปรปรวนของตัวแปรที่จะศึกษาไม่มาก โดยปกติทั่วไป การสุ่มตัวอย่างแบบง่ายสามารถเลือกตัวอย่างได้ 2 ประเภท คือ การสุ่มตัวอย่างแบบแทนที่ และการสุ่มตัวอย่างแบบไม่แทนที่ (ทวีศักดิ์ ศิริพรไพบูลย์, 2549, หน้า 47-52) สำหรับงานวิจัยในครั้งนี้ใช้การสุ่มตัวอย่างแบบไม่แทนที่ และเทคนิคการสุ่มตัวอย่างแบบง่ายอาจทำได้โดย

1.1.1 จับสลาก เป็นเทคนิคการสุ่มตัวอย่างที่เหมาะสำหรับประชากรที่มีขนาดไม่ใหญ่มากนัก โดยมีวิธีคือ เขียนชื่อหรือหมายเลขแทนหน่วยต่าง ๆ ของประชากรลงในฉลากให้ครบ จากนั้นนำใส่กล่อง แล้วจับฉลากที่คละกันนั้นตามจำนวนที่ต้องการ เช่น นักศึกษาห้องหนึ่งมี 30 คน ต้องการให้รางวัลปีใหม่กับนักศึกษาห้องนี้จำนวน 5 รางวัล โดยใช้วิธีการจับฉลาก ถ้าเป็นการสุ่มตัวอย่างแบบใส่คืนจำนวน 5 รางวัล ดังนั้นอาจจะมีนักเรียนบางคนที่ได้รางวัลหลายรางวัล แต่ถ้าเป็นการสุ่มตัวอย่างแบบไม่ใส่คืนจำนวน 5 รางวัล จะมีนักเรียนที่ได้รางวัลดังกล่าวจำนวน 5 คน

1.1.2 ตารางเลขสุ่ม การใช้ตารางเลขสุ่มเป็นเทคนิคการสุ่มตัวอย่างที่เหมาะสำหรับประชากรขนาดใหญ่ และทราบจำนวนที่แน่นอน โดยมีวิธีคือ ให้หมายเลขแก่ทุกหน่วยของประชากร โดยให้จำนวนหลักของหมายเลขเท่ากับหลักของจำนวนประชากร จากนั้นสุ่มหน่วย

ตัวอย่างโดยการดูหมายเลขจากตารางเลขสุ่ม (ตารางเลขสุ่มเป็นตารางที่ประกอบไปด้วย ตัวเลขที่เรียงแบบไม่เจาะจงตามแถว และคอลัมน์ต่าง ๆ) โดยที่จะเริ่มที่แถวใดหรือคอลัมน์ใดก็ได้ แล้วแต่ผู้เลือกเป็นผู้กำหนด อ่านค่าไปจนครบตามจำนวนที่ต้องการ เช่น ถ้าต้องการเลือกตัวอย่างขนาด 10 จากประชากรที่มีขนาด 1,000

ขั้นที่ 1 ให้หมายเลขแก่ประชากร ตั้งแต่ 0001, 0002, . . . , 1000

ขั้นที่ 2 จากนั้นสุ่มเลือกตัวอย่างจากตารางเลขสุ่ม โดยจะเริ่มที่แถวใด คอลัมน์ใดก็ได้ สมมติในที่นี้เริ่มที่แถวที่ 5 คอลัมน์ที่ 1, 2 อ่านค่าจากซ้ายมือไปขวามือ แล้วจึงเริ่มเลข 4 หลักของแถวที่ 6 คอลัมน์ที่ 1, 2 ต่อไป ตามลำดับจนครบตามจำนวนที่ต้องการ คือ 0474, 0163, 0782, 0130, 0647, 0962, 0792, 0676, 0744, 0273

1.1.3 โปรแกรมคอมพิวเตอร์ สำหรับการใช้โปรแกรมคอมพิวเตอร์เพื่อเลือกตัวอย่างสุ่มแบบง่าย จะต้องเริ่มต้นด้วยการกำหนดเลขที่ให้แต่ละหน่วยในประชากร และใช้คำสั่งให้คอมพิวเตอร์เลือกชุดเลขสุ่มเทียมซึ่งจะกำหนดหน่วยตัวอย่างให้ เช่นใช้โปรแกรมสำเร็จรูป SPSS หรือ SAS ในการเลือกตัวอย่างสุ่ม เป็นต้น

1.2 การสุ่มตัวอย่างแบบมีชั้นภูมิ

การสุ่มตัวอย่างแบบมีชั้นภูมิเป็นแผนที่ใช้ข้อมูลเชิงคุณภาพเสริมจากกรอบตัวอย่างเพื่อแบ่งประชากรออกเป็นชั้นภูมิโดยหน่วยตัวอย่างทุกหน่วยของประชากรสามารถจัดอยู่ในชั้นภูมิใดชั้นภูมิหนึ่งได้อย่างชัดเจน และไม่คลุมเครือโดยจัดหน่วยตัวอย่างที่มีคุณลักษณะเหมือนกันไว้ในกลุ่มเดียวกัน และต่างกลุ่มกันจะมีคุณลักษณะที่แตกต่างกันซึ่งการจัดกลุ่มในลักษณะนี้เรียกว่าชั้นภูมิ โดยมีวัตถุประสงค์ เพื่อให้ได้ค่าประมาณที่มีประสิทธิภาพสูงขึ้น และการเลือกตัวอย่างจะเลือกจากแต่ละชั้นภูมิอย่างเป็นอิสระกัน น อกจากนี้เรายังสามารถใช้วิธีการเลือกหน่วยตัวอย่างในแต่ละชั้นภูมิที่แตกต่างกัน เช่นชั้นภูมิที่ 1 อาจใช้การสุ่มตัวอย่างแบบง่าย ชั้นภูมิที่ 2 อาจใช้การสุ่มตัวอย่างแบบมีระบบก็ได้ สมมติ ต้องการศึกษาการจัดการเรียนการสอนของโรงเรียนระดับประถมศึกษาในจังหวัดแห่งหนึ่ง จำนวน 100 โรงเรียน โดยแบ่งประชากรออกเป็น 3 ชั้นภูมิ คือโรงเรียนขนาดใหญ่ 30 โรงเรียน สุ่มตัวอย่างมา 3 โรงเรียนด้วยวิธีการสุ่มตัวอย่างแบบง่าย, โรงเรียนขนาดกลาง 30 โรงเรียน สุ่มตัวอย่างมา 3 โรงเรียนด้วยวิธีการสุ่มตัวอย่างแบบง่าย, และโรงเรียนขนาดเล็ก 40 โรงเรียน สุ่มตัวอย่างมา 4 โรงเรียนด้วยวิธีการสุ่มตัวอย่างแบบมีระบบ เป็นต้น (ทวีศักดิ์ ศิริพรไพบูลย์, 2549, หน้า 96-106)

1.3 การสุ่มตัวอย่างแบบมีระบบ

การสุ่มตัวอย่างแบบมีระบบเป็นวิธีการเลือกตัวอย่างที่มีความสะดวกในการใช้งานหลายกรณีที่การหาหรือสร้างกรอบสำหรับเลือกตัวอย่างกระทำได้ยากหรือไม่สามารถกระทำได้ด้วยเหตุต่าง ๆ เช่นการสำรวจตัวอย่างเกี่ยวกับรถยนต์ที่วิ่งผ่านถนนช่วงหนึ่ง หรือผ่านเมือง ๆ หนึ่ง ซึ่งจะเห็นได้ว่าการหากรอบตัวอย่างของรถยนต์ดังกล่าวเป็นสิ่งที่แทบกระทำไม่ได้เลย เพราะฉะนั้นจึงใช้แผนการสุ่มแบบมีระบบมาแก้ไขปัญหาดังกล่าว และโดยทั่วไปการสุ่มตัวอย่างแบบมีระบบไม่ได้มุ่งหวังในประเด็นเรื่องของคุณภาพเป็นหลัก และไม่คาดหวังให้คุณภาพของตัวประมาณที่ได้สูงกว่าคุณภาพของตัวประมาณจากการเลือกสุ่มตัวอย่างแบบอื่น (สุชาดา กีระนันทน์, 2542, หน้า 103-104)

การสุ่มตัวอย่างแบบมีระบบเป็นวิธีการเลือกหน่วยตัวอย่างตัวแรกด้วยวิธีสุ่ม จากนั้นหน่วยตัวอย่างถัดไปจะได้รับการคัดเลือกอัตโนมัติตามระดับช่วงระยะห่างระหว่างหน่วยตัวอย่างในกรอบตัวอย่างที่กำหนดไว้การสุ่มตัวอย่างแบบมีระบบดำเนินการดังต่อไปนี้

1.3.1 หาช่วงห่างระหว่างสมาชิกที่ถูกสุ่ม (k) โดยนำจำนวนสมาชิกทั้งหมดในกลุ่มประชากรหารด้วยจำนวนสมาชิกในกลุ่มตัวอย่างที่ต้องการสุ่ม ($k = N/n$) เช่น มีสมาชิกในกลุ่มประชากรทั้งหมดจำนวน 500 คน และต้องการกลุ่มตัวอย่างมีขนาด 50 คน ดังนั้น k มีค่าเท่ากับ 10 หรือ $k = 500/50 = 10$

1.3.2 หาตำแหน่งเริ่มของสมาชิกที่ถูกสุ่มโดยผู้วิจัยสุ่มหมายเลขระหว่าง 1 ถึง k ขึ้นมาหมายเลขหนึ่ง หมายเลขนั้นกำหนดให้เป็น r สมมติหมายเลขนั้นคือ 5 ($r = 5$)

1.3.3 สมาชิกหมายเลข r จะได้รับเลือกมาเป็นสมาชิกเริ่มแรกในกลุ่มตัวอย่าง สมาชิกที่ได้รับเลือกต่อไปคือสมาชิกหมายเลข $r+k$, $r+2k$, $r+3k$, ... ตามลำดับจนครบจำนวนที่ต้องการ ถ้าผู้วิจัยสุ่มได้หมายเลข 5 สมาชิกหมายเลข 5 จะได้รับเลือกมาเป็นสมาชิกในกลุ่มตัวอย่าง คนที่ได้รับเลือกต่อไป คือสมาชิกหมายเลข 15, 25, 35 ฯลฯ ตามลำดับจนครบ 50 คน (ทวีศักดิ์ ศิริพรไพบูลย์, 2549, หน้า 126-130)

1.4 การสุ่มตัวอย่างแบบกลุ่ม

การสุ่มตัวอย่างแบบกลุ่มเป็นแผนแบบการสุ่มตัวอย่างที่หน่วยตัวอย่างได้รับการจัดเป็นกลุ่ม โดยจัดให้หน่วยตัวอย่างในกลุ่มเดียวกันมีความแตกต่างหลากหลายคละกัน และในแต่ละกลุ่มหน่วยตัวอย่างที่มีอยู่ในกลุ่มจะไม่มีความแตกต่างกัน โดยทำการเลือกกลุ่มของหน่วยตัวอย่างแทนการเลือกหน่วยตัวอย่างจากประชากรโดยตรง มีวัตถุประสงค์เพื่อความสะดวก หรือลดค่าใช้จ่ายในการจัดเก็บข้อมูล และในกรณีที่จัดเก็บข้อมูลครบทุกหน่วยตัวอย่างในกลุ่มของ

หน่วยตัวอย่างที่สุ่มได้นั้น จะเรียกแผนการสุ่มนี้ว่า แผนการสุ่มตัวอย่างแบบกลุ่มขั้นตอนเดียว (Single-stage Cluster Sampling) เช่น ในการศึกษาปริมาณผลผลิตข้าวของอำเภอหนึ่ง อาจจะเลือกศึกษาเพียงบางหมู่บ้านของอำเภอนั้น แล้วสอบถามปริมาณผลผลิตของทุกครัวเรือนที่ปลูกข้าวในหมูบ้านตัวอย่าง เป็นต้น แต่กระนั้นในบางครั้งคุณลักษณะของหน่วยที่ให้ข้อมูลในกลุ่มตัวอย่างกลุ่มเดียวกันอาจจะมีคุณลักษณะไม่แตกต่างกันมากนักซึ่งอาจไม่จำเป็นต้องศึกษาทุกหน่วยในกลุ่มตัวอย่าง และถ้าจะต้องสุ่มหน่วยที่ให้ข้อมูลบางหน่วยแผนการสุ่มตัวอย่างแบบนี้จะเรียกว่าแผนการสุ่มตัวอย่างแบบกลุ่มสองขั้นตอน (Two-stage Cluster Sampling)โดยขั้นที่ 1 เป็นการสุ่มกลุ่มของหน่วยตัวอย่างที่เป็นกลุ่มตัวอย่างเริ่มต้น และขั้นตอนที่ 2 เป็นการสุ่มตัวอย่างของหน่วยค่าสังเกตซึ่งเป็นหน่วยตัวอย่างในขั้นตอนที่ 2 และในกรณีที่ต้องการสุ่มกลุ่มของหน่วยตัวอย่างมากกว่า 2 ครั้งจะเรียกแผนการสุ่มนี้ว่าแผนการสุ่มตัวอย่างแบบกลุ่มหลายขั้นตอน (Multi-stage Cluster Sampling) โดยขั้นตอนสุดท้ายของการสุ่มตัวอย่างจะต้องเป็นการสุ่มตัวอย่างของหน่วยที่ให้ข้อมูล (ทวีศักดิ์ ศิริพรไพบูลย์, 2549, หน้า 139-151)

2. **แผนการสุ่มตัวอย่างแบบกลุ่มด้วยความน่าจะเป็นไม่เท่ากัน**

ในกระบวนการเลือกหน่วยตัวอย่าง การเลือกตัวอย่างจะใช้ความน่าจะเป็นในการเลือกไม่เท่ากัน และความน่าจะเป็นดังกล่าวจะแฝงอยู่ในขั้นตอนของการเลือกหน่วยตัวอย่าง ดังนั้นโอกาสที่หน่วยตัวอย่างจะได้รับการคัดเลือกสูงมากขึ้นเท่าใด ก็ย่อมแสดงว่าหน่วยตัวอย่างนั้นค่อนข้างสำคัญ และมีผลกระทบต่อการประมาณค่ามากกว่าหน่วยตัวอย่างที่มีโอกาสได้รับการคัดเลือกที่น้อยกว่า เช่น ในการศึกษาเกี่ยวกับความต้องการจ้างลูกจ้างใหม่ในหน่วยงานธุรกิจต่าง ๆ ในเขตภาคเหนือของประเทศไทย ซึ่งเราจะต้องศึกษาจากองค์กรธุรกิจต่าง ๆ ในเขตภาคเหนือบางองค์กรเท่านั้น โดยปกติขนาดของสถานประกอบการใดที่มีขนาดเล็กย่อมมีความต้องการว่าจ้างน้อยกว่าสถานประกอบการที่มีขนาดใหญ่ เพราะฉะนั้นสถานประกอบการที่มีขนาดใหญ่น่าจะมีโอกาสได้รับการคัดเลือกเข้ามาอยู่ในตัวอย่างมากกกว่าสถานประกอบการที่มีขนาดเล็ก เป็นต้น ในกรณีการสุ่มตัวอย่างโดยการใช้ความน่าจะเป็นของการสุ่มไม่เท่ากันสามารถแยกพิจารณาได้เป็น 2 กรณีคือ แบบแทนที่ และไม่แทนที่ ซึ่งในงานวิจัยนี้จะนำเสนอเฉพาะกรณีการใช้ความน่าจะเป็นของการสุ่มไม่เท่ากันแบบไม่แทนที่ (ทวีศักดิ์ ศิริพรไพบูลย์, 2549, หน้า 65)

ทฤษฎี 1 วิธีการความน่าจะเป็นไม่เท่ากันตามสัดส่วนกับขนาดของกลุ่มหน่วยตัวอย่าง

การเลือกตัวอย่างจากแผนการสุ่มตัวอย่างแบบกลุ่มขั้นตอนเดียวไม่แทนที่จำนวน n กลุ่มจากจำนวนกลุ่มตัวอย่างประชากรทั้งสิ้น N กลุ่ม ด้วยวิธีการความน่าจะเป็นไม่เท่ากันตามสัดส่วนกับขนาดของกลุ่มหน่วยตัวอย่าง

1. ตัวประมาณที่ไม่เอนเอียงของค่าเฉลี่ยตัวอย่าง

$$\bar{y}_{CLHT} = \frac{1}{N}\sum_{i=1}^{v}\frac{\tau_i}{\pi_i} \text{ โดยที่ } \tau_i = \sum y_{ij} \text{ และ}$$

v เป็นจำนวนกลุ่มหน่วยตัวอย่างที่แตกต่างกันทั้งหมดจากการสุ่ม n ครั้ง

π_i คือความน่าจะเป็นที่กลุ่มหน่วยตัวอย่างที่ i จะอยู่ในชุดตัวอย่าง ซึ่งมีค่าเท่ากับ $\pi_i \approx 1-(1-p_i)^n$

2. ความแปรปรวนของค่าประมาณค่าเฉลี่ยตัวอย่าง

$$V(\bar{y}_{CLHT}) = \frac{1}{N^2}\left[\sum_{i=1}^{N}\left(\frac{1-\pi_i}{\pi_i}\right)\tau_i^2 + \sum_{i=1}^{N}\sum_{j\neq i}^{N}\left(\frac{\pi_{ij}-\pi_i\pi_j}{\pi_i\pi_j}\right)\tau_i\tau_j\right]$$

โดยที่ π_{ij} คือความน่าจะเป็นที่กลุ่มหน่วยตัวอย่างที่ i และ j จะอยู่ในชุดตัวอย่าง ซึ่งมีค่าเท่ากับ $\pi_{ij} \approx \pi_i + \pi_j - \left[1-\left(1-p_i-p_j\right)^n\right]$

3. ค่าประมาณความแปรปรวนของค่าประมาณค่าเฉลี่ยตัวอย่าง

$$v\left(\bar{y}_{CLHT}\right) = \frac{1}{N^2}\left[\sum_{i=1}^{v}\left(\frac{1-\pi_i}{\pi_i^2}\right)\tau_i^2 + \sum_{i=1}^{v}\sum_{j\neq i}^{v}\left(\frac{\pi_{ij}-\pi_i\pi_j}{\pi_i\pi_j}\right)\frac{\tau_i\tau_j}{\pi_{ij}}\right]$$

พิสูจน์

1. กำหนดให้ $l_i,\ i=1,2,\ldots,N$ เป็นตัวแปรสุ่มที่มีค่าเท่ากับ 1 ถ้าหน่วยตัวอย่างที่ i ตกอยู่ในชุดตัวอย่าง และมีค่าเท่ากับ 0 ถ้าหน่วยตัวอย่างที่ i ไม่ได้ตกอยู่ในชุดตัวอย่าง ดังนั้น l_i จะมีการแจกแจงแบบเบอร์นูลีด้วยความน่าจะเป็นเท่ากับ π_i หรือ $l_i \sim Ber(\pi_i)$ ดังนั้น

$$P(l_i=1)=\pi_i\ ,\ P(l_i=0)=1-\pi_i$$

และ

$$E(l_i) = 1.P(l_i = 1) + 0.P(l_i = 0) = \pi_i$$

ในทำนองเดียวกันจะได้

$$E(l_j) = \pi_j$$

เพราะฉะนั้น

$$E(\overline{y}_{CLHT}) = E\left(\frac{1}{N}\sum_{i=1}^{N}\frac{l_i\tau_i}{\pi_i}\right)$$

$$= \frac{1}{N}\sum_{i=1}^{N}\frac{\tau_i E(l_i)}{\pi_i}$$

$$= \frac{1}{N}\sum_{i=1}^{N}\frac{\tau_i\pi_i}{\pi_i}$$

$$= \frac{1}{N}\sum_{i=1}^{N}\tau_i$$

$$= \mu$$

2. นอกจากนี้ l_i ยังมีคุณสมบัติดังนี้

$$V(l_i) = E(l_i^2) - [E(l_i)]^2$$

$$= E(l_i) - [E(l_i)]^2$$

$$= \pi_i - \pi_i^2$$

$$= \pi_i(1 - \pi_i)$$

$$E(l_i l_j) = \sum_{\text{all i,all j}} l_i l_j p(l_i l_j)$$

$$= 0(0)p(l_i = 0, l_j = 0) + 1(0)p(l_i = 1, l_j = 0) +$$

$$+0(1)p(l_i = 0, l_j = 0) + 1(1)p(l_i = 1, l_j = 1)$$

$$= p(l_i = 1, l_j = 1)$$

$$= \pi_{ij}$$

และ

$$Cov(l_{i,} l_j) = E(l_i l_j) - E(l_i)E(l_j)$$

$$= \pi_{ij} - \pi_i \pi_j$$

กำหนดให้ $Z_i = \dfrac{\tau_i}{\pi_i}$ และ $Z_j = \dfrac{\tau_j}{\pi_j}$ เพื่อใช้ในการพิสูจน์ $V(\bar{y}_{CLHT})$ จะได้ว่า

$$V(\bar{y}_{CLHT}) = V\left(\frac{1}{N}\sum_{i=1}^{N} Z_i l_i\right)$$

$$= \frac{1}{N^2} V\left(\sum_{i=1}^{N} l_i Z_i\right)$$

$$= \frac{1}{N^2}\left[\sum_{i=1}^{N}\left(\frac{\tau_i}{\pi_i}\right)^2 V(l_i) + \sum_{i=1}^{N}\sum_{j\neq 1}^{N}\left(\frac{\tau_i \tau_j}{\pi_i \pi_j}\right)\text{cov}(l_i l_j)\right]$$

$$= \frac{1}{N^2}\left[\sum_{i=1}^{N}\left(\frac{\tau_i}{\pi_i}\right)^2 \pi_i(1-\pi_i) + \sum_{i=1}^{N}\sum_{j\neq i}^{N}\left(\frac{\tau_i \tau_j}{\pi_i \pi_j}\right)\pi_{ij} - \pi_i \pi_j\right]$$

$$= \frac{1}{N^2}\left[\sum_{i=1}^{N}\left(\frac{1-\pi_i}{\pi_i}\right)\tau_i^2 + \sum_{i=1}^{N}\sum_{j\neq i}^{N}\left(\frac{\pi_{ij}-\pi_i\pi_j}{\pi_i\pi_j}\right)\tau_i\tau_j\right]$$

3. ค่าประมาณความแปรปรวนของ $v(\bar{y}_{CLHT})$ เป็นดังนี้

$$v(\bar{y}_{CLHT}) = \frac{1}{N^2}\left[\sum_{i=1}^{v}\left(\frac{1-\pi_i}{\pi_i^2}\right)\tau_i^2 + \sum_{i=1}^{v}\sum_{j\neq i}^{v}\left(\frac{\pi_{ij}-\pi_i\pi_j}{\pi_i\pi_j}\right)\frac{\tau_i\tau_j}{\pi_{ij}}\right]$$

ดังนั้น

$$E\left(v(\bar{y}_{CLHT})\right) = E\left\{\frac{1}{N^2}\left[\sum_{i=1}^{N}\left(\frac{1-\pi_i}{\pi_i^2}\right)\tau_i^2 l_i + \sum_{i=1}^{N}\sum_{j\neq i}^{N}\left(\frac{\pi_{ij}-\pi_i\pi_j}{\pi_i\pi_j}\right)\frac{\tau_i\tau_j l_i l_j}{\pi_{ij}}\right]\right\}$$

$$= \frac{1}{N^2}\left[\sum_{i=1}^{N}\left(\frac{1-\pi_i}{\pi_i^2}\right)\tau_i^2 E(l_i) + \sum_{i=1}^{N}\sum_{j\neq i}^{N}\left(\frac{\pi_{ij}-\pi_i\pi_j}{\pi_i\pi_j}\right)\frac{\tau_i\tau_j E(l_i l_j)}{\pi_{ij}}\right] ; E(l_i l_j) = \pi_{ij}$$

$$= \frac{1}{N^2}\left[\sum_{i=1}^{N}\left(\frac{1-\pi_i}{\pi_i^2}\right)\tau_i^2 \pi_i + \sum_{i=1}^{N}\sum_{j\neq i}^{N}\left(\frac{\pi_{ij}-\pi_i\pi_j}{\pi_i\pi_j}\right)\frac{\tau_i\tau_j \pi_{ij}}{\pi_{ij}}\right]$$

$$= \frac{1}{N^2}\left[\sum_{i=1}^{N}\left(\frac{1-\pi_i}{\pi_i}\right)\tau_i^2 + \sum_{i=1}^{N}\sum_{j\neq i}^{N}\left(\frac{\pi_{ij}-\pi_i\pi_j}{\pi_i\pi_j}\right)\tau_i\tau_j\right]$$

$$= V(\bar{y}_{CLHT})$$

3. วิธีการประมาณค่าเฉลี่ยด้วยความน่าจะเป็นเท่ากัน

โดยปกติการประมาณค่าเฉลี่ยเราจะใช้วิธีการประมาณค่า 2 วิธี ดังนี้

3.1 การประมาณค่าเฉลี่ยแบบง่าย

การประมาณค่าเฉลี่ยแบบง่ายเป็นการประมาณค่าโดยอาศัยข้อมูลที่เราสนใจเท่านั้น เช่นการประมาณค่าเฉลี่ย $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ หรือการประมาณค่าผลรวม $\hat{\tau} = N\bar{y}$ ซึ่งจะเห็นได้ว่าเราจะใช้ข้อมูลเฉพาะตัวแปรที่เราสนใจศึกษาเท่านั้น (ทวีศักดิ์ ศิริพรไพบูลย์, 2549, หน้า 47-49) ในงานวิจัยนี้กล่าวเฉพาะกรณีที่เลือกหน่วยตัวอย่างด้วยความน่าจะเป็นที่เท่ากันแบบไม่แทนที่

ทฤษฎี 2 การประมาณค่าเฉลี่ยแบบไม่แทนที่อย่างง่าย

การสุ่มตัวอย่างแบบไม่แทนที่อย่างง่ายขนาด n หน่วยจากประชากรขนาด N หน่วย

1. ตัวประมาณที่ไม่เอนเอียงของค่าเฉลี่ยตัวอย่าง

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

2. ความแปรปรวนของค่าประมาณค่าเฉลี่ย

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

3. ค่าประมาณความแปรปรวนของค่าประมาณค่าเฉลี่ย

$$v(\bar{y}) = \frac{N-n}{N}\frac{s^2}{n} \text{ โดยที่ } s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$$

พิสูจน์

1. เนื่องจาก $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \; = \frac{1}{n}\sum_{i=1}^{N} a_i y_i ; a_i \sim Ber\left(\frac{1}{N}\right)$ และ

$$p(a_i = 1) = \frac{\text{\# sample including unit i}}{\text{\# possible samples}}$$

$$= \binom{N-1}{n-1} \Big/ \binom{N}{n}$$

$$= \frac{n}{N}$$

และ

$$E(a_i) = \sum_{\text{all i}} a_i p(a_i)$$

$$= 0.P(a_i = 0) + 1.P(a_i = 1)$$

$$= \frac{n}{N}$$

ดังนั้น

$$E(\bar{y}) = E\left[\frac{1}{n}\sum_{i=1}^{N} a_i y_i\right]$$

$$= \frac{1}{n}\sum_{i=1}^{N} E(a_i y_i)$$

$$= \frac{1}{n}\sum_{i=1}^{N} y_i E(a_i)$$

$$= \frac{1}{n}\sum_{i=1}^{N} y_i \left(\frac{n}{N}\right)$$

$$= \frac{1}{N}\sum_{i=1}^{N} y_i$$

$$= \mu$$

2. สำหรับ $V(a_i) = E(a_i^2) - [E(a_i)]^2$ เนื่องจาก

$$E\left(a_i^2\right) = 0^2\, p\left(a_i = 0\right) + 1^2\, p\left(a_i = 1\right)$$

$$= \frac{n}{N}$$

ดังนั้น

$$V(a_i) = \frac{n}{N} - \left(\frac{n}{N}\right)^2$$

$$= \frac{n}{N}\left(1 - \frac{n}{N}\right)$$

สำหรับ $i \neq j$

$$E(a_i a_j) = 0(0)p(a_i = 0, a_j = 0) + 0(1)p(a_i = 0, a_j = 1)$$

$$+1(0)p(a_i = 1, a_j = 0) + 1(1)p(a_i = 1, a_j = 1)$$

$$= p(a_i = 1, a_j = 1)$$

$$= p(a_i = 1)p(a_j = 1 \mid a_i = 1)$$

$$= \frac{n}{N}\left(\frac{n-1}{N-1}\right)$$

ดังนั้น

$$Cov(a_i, a_j) = E(a_i a_j) - E(a_i)E(a_j)$$

$$= \frac{n}{N}\left(\frac{n-1}{N-1}\right) - \left(\frac{n}{N}\cdot\frac{n}{N}\right)$$

$$= \frac{n}{N}\left(\frac{n-1}{N-1} - \frac{n}{N}\right)$$

$$= \frac{n}{N}\left(\frac{nN - N - nN + n}{N(N-1)}\right)$$

$$= -\frac{n}{N}\left(\frac{N-n}{N(N-1)}\right)$$

$$= -\frac{1}{N-1}\left(\frac{N-n}{N}\right)\left(\frac{n}{N}\right)$$

$$= -\frac{1}{N-1}\left(\frac{n}{N}\right)\left(1 - \frac{n}{N}\right)$$

เพราะฉะนั้น

$$V(\bar{y}) = \frac{1}{n^2} V\left(\sum_{i=1}^{n} y_i\right)$$

$$= \frac{1}{n^2} V\left(\sum_{i=1}^{N} a_i y_i\right)$$

$$= \frac{1}{n^2} Cov\left(\sum_{i=1}^{N} a_i y_i, \sum_{j=1}^{N} a_j y_j\right)$$

$$= \frac{1}{n^2}\left[\sum_{i=1}^{N} y_i^2 V(a_i) + \sum_{i=1}^{N}\sum_{j\neq i}^{N} y_i y_j Cov(a_i, a_j)\right]$$

$$= \frac{1}{n^2}\left[\frac{n}{N}\left(1-\frac{n}{N}\right)\sum_{i=1}^{N} y_i^2 - \sum_{i=1}^{N}\sum_{j\neq i}^{N} y_i y_j \frac{1}{N-1}\left(\frac{n}{N}\right)\left(1-\frac{n}{N}\right)\right]$$

$$= \frac{1}{n^2}\left(\frac{n}{N}\right)\left(1-\frac{n}{N}\right)\left[\sum_{i=1}^{N} y_i^2 - \frac{1}{N-1}\sum_{i=1}^{N}\sum_{j\neq i}^{N} y_i y_j\right]$$

$$= \frac{1}{n}\left(1-\frac{n}{N}\right)\left(\frac{1}{N(N-1)}\right)\left[(N-1)\sum_{i=1}^{N} y_i^2 - \left(\sum_{i=1}^{N} y_i\right)^2 + \sum_{i=1}^{N} y_i^2\right]$$

$$= \frac{1}{n}\left(1-\frac{n}{N}\right)\left(\frac{1}{N(N-1)}\right)\left[N\sum_{i=1}^{N} y_i^2 - \left(\sum_{i=1}^{N} y_i\right)^2\right]$$

$$= \frac{1}{n}\left(1-\frac{n}{N}\right)\left(\frac{1}{(N-1)}\right)\left[\sum_{i=1}^{N} y_i^2 - \frac{1}{N}\left(\sum_{i=1}^{N} y_i\right)^2\right]$$

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

$$= \left(\frac{N-n}{N}\right)\frac{S^2}{n}$$

3. เนื่องจาก

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(y_i - \mu + \mu - \bar{y})^2$$

$$= \sum_{i=1}^{n}(y_i - \mu)^2 - n(\bar{y} - \mu)^2$$

ดังนั้น

$$E\left[\sum_{i=1}^{n}(y_i - \bar{y})^2\right] = E\left[\sum_{i=1}^{n}(y_i - \mu)^2\right] - nV(\bar{y})$$

$$= E\left[\sum_{i=1}^{N}a_i(y_i - \mu)^2\right] - nV(\bar{y})$$

$$= \frac{n}{N}\sum_{i=1}^{N}(y_i - \mu)^2 - \left(\frac{N-n}{N}\right)S^2$$

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

$$= (n-1)S^2$$

และ

$$E(v(\bar{y})) = \left(1 - \frac{n}{N}\right)\frac{1}{n}E(s^2)$$

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

$$= V(\bar{y})$$

3.2 การประมาณค่าเฉลี่ยแบบอัตราส่วน

เป็นการประมาณค่าเฉลี่ยประชากร μ ของตัวแปรที่ศึกษา (Y) โดยใช้ประโยชน์จากตัวแปรช่วย (X) ที่มีความสัมพันธ์กันเชิงเส้นตรงกับตัวแปรที่ศึกษา ดังทฤษฎี 3 (ทวีศักดิ์ ศิริพรไพบูลย์, 2549, หน้า 74-78)

ทฤษฎี 3 การประมาณค่าเฉลี่ยแบบอัตราส่วนอย่างง่าย

การสุ่มตัวอย่างแบบง่ายไม่แทนที่จำนวน n หน่วย จากประชากรขนาด N หน่วย เมื่อตัวแปรที่ศึกษา และตัวแปรช่วยมีความสัมพันธ์กันเชิงเส้นตรง

1. ตัวประมาณค่าเฉลี่ยแบบอัตราส่วนตัวอย่าง $\bar{y}_R$ เป็นตัวประมาณที่เอนเอียงสำหรับค่าเฉลี่ยประชากร μ มีค่าเท่ากับ

$$\bar{y}_R = \hat{R}\mu_x$$

เมื่อ

$$\hat{R} = \frac{\bar{y}}{\bar{x}} \text{ และ } \hat{\mu}_x = \frac{1}{n}\sum_{i=1}^{n} x_i$$

2. ความแปรปรวนของตัวประมาณค่าเฉลี่ยอัตราส่วนตัวอย่างมีค่าเท่ากับ

$$V(\bar{y}_R) = \left(1 - \frac{n}{N}\right)\frac{S_d^2}{n}$$

เมื่อ

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

3. ค่าประมาณความแปรปรวนของตัวประมาณอัตราส่วนค่าเฉลี่ยตัวอย่างมีค่าเท่ากับ

$$v(\bar{y}_R) = \left(1 - \frac{n}{N}\right)\frac{s_d^2}{n}$$

เมื่อ

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

พิสูจน์

1. กำหนดให้ $\hat{R}=\dfrac{\bar{y}}{\bar{x}}$ ซึ่งจะเห็นได้ว่า $\hat{R}$ คือค่าเฉลี่ย $\bar{y}$ หารด้วย $\bar{x}$ โดยที่การแจกแจงสิ่งตัวอย่างของ $\hat{R}$ จะมีลักษณะเบ้ และค่อนข้างยุ่งยากมากกว่าค่าเฉลี่ย $\bar{y}$ เนื่องจากทั้งตัวตั้ง $\bar{y}$ และตัวหาร $\bar{x}$ จะเปลี่ยนแปลงจากตัวอย่างชุดหนึ่งไปยังอีกตัวอย่างชุดหนึ่งโดยเฉพาะขนาดตัวอย่างน้อย ดังนั้นโดยปกติ $\hat{R}$ จะเป็นตัวประมาณที่เอนเอียงสำหรับ R เนื่องจาก

$$\bar{x}=\mu_x+(\bar{x}-\mu_x)=\mu_x\left(1+\frac{\bar{x}-\mu_x}{\mu_x}\right) \text{ และ } \bar{y}=\mu+(\bar{y}-\mu)=\mu\left(1+\frac{\bar{y}-\mu}{\mu}\right)$$

ดังนั้น

$$\hat{R}=\frac{\bar{y}}{\bar{x}}=\frac{\mu}{\mu_x}\left(1+\frac{\bar{y}-\mu}{\mu}\right)\left(1+\frac{\bar{x}-\mu_x}{\mu_x}\right)^{-1}$$

$$=\frac{\mu}{\mu_x}\left(1+\frac{\bar{y}-\mu}{\mu}\right)\left(1-\frac{\bar{x}-\mu_x}{\mu_x}+\frac{(\bar{x}-\mu_x)^2}{\mu_x^2}+\ldots\right)$$

$$=R\left(1-\frac{\bar{x}-\mu_x}{\mu_x}+\frac{\bar{y}-\mu}{\mu}+\frac{(\bar{x}-\mu_x)^2}{\mu_x^2}-\frac{(\bar{x}-\mu_x)}{\mu_x}\frac{(\bar{y}-\mu)}{\mu}+\ldots\right)$$

เนื่องจาก $E(\hat{R})\neq R$ ดังนั้น $\hat{R}$ เป็นตัวประมาณที่เอนเอียงของ R

สำหรับตัวประมาณค่าเฉลี่ยแบบอัตราส่วน $\bar{y}_R=\hat{R}\mu_x$ จะได้ว่า

$$E(\bar{y}_R)=E(\hat{R}\mu_x)$$

$$=\mu_x E(\hat{R})$$

$$\neq \mu_y$$

2. เนื่องจาก $\hat{R}$ เป็นตัวประมาณที่เอนเอียงสำหรับ R ดังนั้น

$$MSE(\hat{R})=E(\hat{R}-R)^2$$

$$= E\left(\frac{\bar{y}}{\bar{x}} - R\right)^2$$

$$= E\left(\frac{\bar{y} - R\bar{x}}{\bar{x}}\right)^2$$

เนื่องจาก $\bar{y}$ และ $\bar{x}$ ไม่ใช่ค่าคงที่ จะประมาณ $\bar{x}$ ด้วย μ_x ดังนั้น

$$MSE\left(\hat{R}\right) \cong \frac{1}{\mu_x^2} E(\bar{y} - R\bar{x})^2$$

กำหนดให้ $d_i = y_i - Rx_i$ ดังนั้น $\bar{y}_d = \frac{1}{n}\sum_{i-1}^{n} d_i = \bar{y} - R\bar{x}$ และ

$$E(\bar{y}_d) = E(\bar{y} - R\bar{x}) = \mu - R\mu_x = \mu - \frac{\mu}{\mu_x}\mu_x = 0$$

ดังนั้น

$$MSE\left(\hat{R}\right) \cong V(\hat{R}) = \frac{1}{\mu_x^2} E\left(\bar{y}_d^2\right)$$

$$= \frac{1}{\mu_x^2} V\left(\bar{y}_d\right)$$

$$= \frac{1}{\mu_x^2}\left(1 - \frac{n}{N}\right)\frac{S_d^2}{n}$$

$$= \frac{1}{\mu_x^2}\left(1 - \frac{n}{N}\right)\frac{1}{n}\frac{1}{N-1}\sum_{i=1}^{N}(d_i - E(\bar{y}_d))^2$$

$$= \frac{1}{n\mu_x^2}\left(1 - \frac{n}{N}\right)\frac{1}{N-1}\sum_{i=1}^{N}(y_i - Rx_i)^2$$

$$= \frac{1}{\mu_x^2}\left(1 - \frac{n}{N}\right)\frac{S_d^2}{n}$$

เมื่อ

$$S_d^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(y_i - Rx_i\right)^2$$

ดังนั้นความแปรปรวนของตัวประมาณค่าเฉลี่ยอัตราส่วนประมาณได้

$$V(\bar{y}_R) = \left(1 - \frac{n}{N}\right)\frac{S_d^2}{n}$$

3. เนื่องจาก $\sum_{i=1}^{n}(y_i - Rx_i)^2 = \sum_{i=1}^{N} a_i (y_i - Rx_i)^2$

$$E\left[\sum_{i=1}^{n}(y_i - Rx_i)^2\right] = \sum_{i=1}^{N}(y_i - Rx_i)^2 E(a_i)$$

$$= \frac{n}{N}\sum_{i=1}^{N}(y_i - Rx_i)^2$$

และ

$$E(s_d^2) = E\left[\frac{1}{n-1}\sum_{i=1}^{n}(y_i - Rx_i)^2\right]$$

$$= \frac{n}{N(n-1)}\sum_{i=1}^{N}(y_i - Rx_i)^2$$

$$= \frac{n}{(n-1)}\frac{N-1}{N}S_d^2$$

ดังนั้น

$$E\left(\frac{1}{n\mu_x^2}\left(1 - \frac{n}{N}\right)s_d^2\right) = \frac{1}{n\mu_x^2}\left(1 - \frac{n}{N}\right)\frac{n}{n-1}\frac{N-1}{N}S_d^2$$

$$= \frac{n}{n-1}\frac{1}{\mu_x^2}\left(1 - \frac{n}{N}\right)\left(\frac{N-1}{N}\right)\frac{S_d^2}{n}$$

และค่าประมาณความแปรปรวนของอัตราส่วนตัวอย่างมีค่าเท่ากับ

$$v\left(\hat{R}\right)=\frac{1}{\mu_x^2}\left(1-\frac{n}{N}\right)\left(\frac{N-1}{N}\right)\left(\frac{n}{n-1}\right)\frac{s_d^2}{n}$$

หรือ

$$v\left(\hat{R}\right)\cong\frac{1}{\mu_x^2}\left(1-\frac{n}{N}\right)\frac{s_d^2}{n}$$

เมื่อ

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

ดังนั้นค่าประมาณความแปรปรวนของตัวประมาณค่าเฉลี่ยอัตราส่วนตัวอย่าง มีค่าเท่ากับ

$$v\left(\bar{y}_R\right)=\left(1-\frac{n}{N}\right)\frac{s_d^2}{n}$$

4. วิธีการประมาณค่าเฉลี่ยด้วยความน่าจะเป็นไม่เท่ากัน

วิธีการประมาณค่าโดยใช้ความน่าจะเป็นไม่เท่ากัน เป็นวิธีการที่เลือกหน่วยตัวอย่างใดเข้าไปในชุดตัวอย่างแล้วจะไม่ได้รับการคัดเลือกหน่วยตัวอย่างนั้นในครั้งถัดไปของการคัดเลือก ดังนั้นในการคัดเลือกหน่วยตัวอย่างแต่ละครั้งความน่าจะเป็นของหน่วยตัวอย่างที่จะได้รับการคัดเลือกจะเป็นไปตามเงื่อนไขของหน่วยที่ได้รับการคัดเลือกเข้าไปในชุดตัวอย่างก่อนหน้า ซึ่งมีวิธีการประมาณค่า 2 วิธีดังนี้

4.1 การประมาณค่าเฉลี่ยอย่างง่าย

ในการประมาณค่าเฉลี่ยอย่างง่ายด้วยวิธีการความน่าจะเป็นไม่เท่ากันจากการเลือกหน่วยตัวอย่างแบบไม่แทนที่ ของ Horvitz-Thompson (1952) ดังทฤษฎี 4 (ทวีศักดิ์ ศิริพรไพบูลย์, 2549, หน้า 66)

ทฤษฎี 4 การประมาณค่าเฉลี่ยด้วยวิธีการของ Horvitz - Thompson

กำหนดให้ y_i เป็นค่าของ y ในหน่วยตัวอย่างที่เลือกมาในครั้งที่ i และ p_i จะเป็นความน่าจะเป็นที่เลือกหน่วยตัวอย่างในครั้งที่ i , $i=1,\ldots,N$

1. ตัวประมาณที่ไม่เอนเอียงของค่าเฉลี่ยตัวอย่าง

$$\bar{y}_{HT} = \frac{1}{N}\sum_{i=1}^{n}\frac{y_i}{\pi_i}$$

π_i คือความน่าจะเป็นที่กลุ่มหน่วยตัวอย่างที่ i จะอยู่ในชุดตัวอย่าง ซึ่งมีค่าเท่ากับ $\pi_i \approx 1-(1-p_i)^n$

2. ความแปรปรวนของค่าประมาณค่าเฉลี่ยตัวอย่าง

$$V(\bar{y}_{HT}) = \frac{1}{N^2}\left[\sum_{i=1}^{N}\left(\frac{1-\pi_i}{\pi_i}\right)y_i^2 + \sum_{i=1}^{N}\sum_{j\neq i}^{N}\left(\frac{\pi_{ij}-\pi_i\pi_j}{\pi_i\pi_j}\right)y_i y_j\right]$$

โดยที่ π_{ij} คือความน่าจะเป็นที่กลุ่มหน่วยตัวอย่างที่ i และ j จะอยู่ในชุดตัวอย่าง ซึ่งมีค่าเท่ากับ $\pi_{ij} \approx \pi_i + \pi_j - \left[1-\left(1-p_i-p_j\right)^n\right]$

3. ค่าประมาณความแปรปรวนของค่าประมาณค่าเฉลี่ยตัวอย่าง

$$v\left(\bar{y}_{HT}\right) = \frac{1}{N^2}\left[\sum_{i=1}^{n}\left(\frac{1-\pi_i}{\pi_i^2}\right)y_i^2 + \sum_{i=1}^{n}\sum_{j\neq i}^{n}\left(\frac{\pi_{ij}-\pi_i\pi_j}{\pi_i\pi_j}\right)\frac{y_i y_j}{\pi_{ij}}\right]$$

พิสูจน์

1. กำหนดให้ $l_i,\ i=1,2,...,N$ เป็นตัวแปรสุ่มที่มีค่าเท่ากับ 1 ถ้าหน่วยตัวอย่างที่ i ตกอยู่ในชุดตัวอย่าง และมีค่าเท่ากับ 0 ถ้าหน่วยตัวอย่างที่ i ไม่ได้ตกอยู่ในชุดตัวอย่าง ดังนั้น l_i จะมีการแจกแจงแบบเบอร์นูลีด้วยความน่าจะเป็นเท่ากับ π_i หรือ $l_i \sim Ber(\pi_i)$ ดังนั้น

$$P(l_i=1)=\pi_i\,,\ P(l_i=0)=1-\pi_i$$

และ

$$E(l_i) = 1.P(l_i=1)+0.P(l_i=0) = \pi_i$$

ในทำนองเดียวกันจะได้

$$E(l_j) = \pi_j$$

เพราะฉะนั้น

$$E(\bar{y}_{HT}) = E\left(\frac{1}{N}\sum_{i=1}^{N}\frac{l_i y_i}{\pi_i}\right)$$

$$= \frac{1}{N}\sum_{i=1}^{N}\frac{y_i \pi_i}{\pi_i}$$

$$= \frac{1}{N}\sum_{i=1}^{N} y_i$$

$$= \mu$$

2. นอกจากนี้ l_i ยังมีคุณสมบัติดังนี้

$$V(l_i) = E(l_i^2) - [E(l_i)]^2$$

$$= E(l_i) - [E(l_i)]^2$$

$$= \pi_i - \pi_i^2$$

$$= \pi_i(1-\pi_i)$$

$$E(l_i l_j) = \sum_{\text{all i,all j}} l_i l_j p(l_i l_j)$$

$$= 0(0)p(l_i = 0, l_j = 0) + 1(0)p(l_i = 1, l_j = 0) +$$

$$+0(1)p(l_i = 0, l_j = 0) + 1(1)p(l_i = 1, l_j = 1)$$

$$= p(l_i = 1, l_j = 1)$$

$$= \pi_{ij}$$

และ

$$Cov(l_i l_j) = E(l_i l_j) - E(l_i)E(l_j)$$

$$= \pi_{ij} - \pi_i \pi_j$$

-กำหนดให้ $Z_i = \dfrac{y_i}{\pi_i}$ และ $Z_j = \dfrac{y_j}{\pi_j}$ เพื่อใช้ในการพิสูจน์ $V(\bar{y}_{HT})$ จะได้ว่า

$$V(\bar{y}_{HT}) = V\left(\frac{1}{N}\sum_{i=1}^{N} Z_i l_i\right)$$

$$= \frac{1}{N^2} V\left(\sum_{i=1}^{N} l_i Z_i\right)$$

$$= \frac{1}{N^2}\left[\sum_{i=1}^{N}\left(\frac{y_i}{\pi_i}\right)^2 V(l_i) + \sum_{i=1}^{N}\sum_{j\neq 1}^{N}\left(\frac{y_i y_j}{\pi_i \pi_j}\right)\mathrm{cov}(l_i l_j)\right]$$

$$= \frac{1}{N^2}\left[\sum_{i=1}^{N}\left(\frac{y_i}{\pi_i}\right)^2 \pi_i(1-\pi_i) + \sum_{i=1}^{N}\sum_{j\neq i}^{N}\left(\frac{y_i y_j}{\pi_i \pi_j}\right)\pi_{ij} - \pi_i \pi_j\right]$$

$$= \frac{1}{N^2}\left[\sum_{i=1}^{N}\left(\frac{1-\pi_i}{\pi_i}\right)y_i^2 + \sum_{i=1}^{N}\sum_{j\neq i}^{N}\left(\frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j}\right)y_i y_j\right]$$

3. ค่าประมาณความแปรปรวนของ $v(\bar{y}_{HT})$ เป็นดังนี้

$$v(\bar{y}_{HT}) = \frac{1}{N^2}\left[\sum_{i=1}^{n}\left(\frac{1-\pi_i}{\pi_i^2}\right)y_i^2 + \sum_{i=1}^{n}\sum_{j\neq i}^{n}\left(\frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j}\right)\frac{y_i y_j}{\pi_{ij}}\right]$$

ดังนั้น

$$E\left(v(\bar{y}_{HT})\right) = E\left\{\frac{1}{N^2}\left[\sum_{i=1}^{N}\left(\frac{1-\pi_i}{\pi_i^2}\right)y_i^2 l_i + \sum_{i=1}^{N}\sum_{j\neq i}^{N}\left(\frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j}\right)\frac{y_i y_j l_i l_j}{\pi_{ij}}\right]\right\}$$

$$= \frac{1}{N^2}\left[\sum_{i=1}^{N}\left(\frac{1-\pi_i}{\pi_i^2}\right)y_i^2 E(l_i) + \sum_{i=1}^{N}\sum_{j\neq i}^{N}\left(\frac{\pi_{ij}-\pi_i\pi_j}{\pi_i\pi_j}\right)\frac{y_i y_j E(l_i l_j)}{\pi_{ij}}\right];E(l_i l_j)=\pi_{ij}$$

$$= \frac{1}{N^2}\left[\sum_{i=1}^{N}\left(\frac{1-\pi_i}{\pi_i^2}\right)y_i^2 \pi_i + \sum_{i=1}^{N}\sum_{j\neq i}^{N}\left(\frac{\pi_{ij}-\pi_i\pi_j}{\pi_i\pi_j}\right)\frac{y_i y_j \pi_{ij}}{\pi_{ij}}\right]$$

$$= \frac{1}{N^2}\left[\sum_{i=1}^{N}\left(\frac{1-\pi_i}{\pi_i}\right)y_i^2 + \sum_{i=1}^{N}\sum_{j\neq i}^{N}\left(\frac{\pi_{ij}-\pi_i\pi_j}{\pi_i\pi_j}\right)y_i y_j\right]$$

$$= V(\bar{y}_{HT})$$

4.2 การประมาณค่าเฉลี่ยแบบอัตราส่วน

Thompson (2002) ได้ศึกษาวิธีการประมาณค่าเฉลี่ยแบบอัตราส่วน และความแปรปรวนของตัวประมาณค่าเฉลี่ยแบบอัตราส่วนภายใต้แผนการสุ่มตัวอย่างแบบง่ายโดยมีรายละเอียดดังทฤษฎี 5

ทฤษฎี 5 การประมาณค่าเฉลี่ยแบบอัตราส่วน

การสุ่มตัวอย่างแบบง่ายไม่แทนที่จำนวน n หน่วย จากประชากรขนาด N หน่วย ตัวแปรที่สนใจศึกษาคือ Y และตัวแปรช่วยเชิงปริมาณที่สัมพันธ์กับ Y เชิงเส้นตรงคือ ตัวแปร X

1. ตัวประมาณค่าเฉลี่ยแบบอัตราส่วนที่เอนเอียงสำหรับค่าเฉลี่ยประชากร μ เท่ากับ

$$\bar{y}_{rHT} = \frac{\bar{y}_{HT}}{\bar{x}_{HT}}\mu_x = \hat{R}_{HT}\mu_x$$

โดยที่ $\bar{y}_{HT} = \frac{1}{N}\sum_{i=1}^{n}\frac{y_i}{\pi_i}$, $\bar{x}_{HT} = \frac{1}{N}\sum_{i=1}^{n}\frac{x_i}{\pi_i}$, $\mu_x = \frac{1}{N}\sum_{i=1}^{N}x_i$ และ

2. ความแปรปรวนของตัวประมาณค่าเฉลี่ยแบบอัตราส่วนเท่ากับ

$$V(\bar{y}_{rHT}) = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}y_i' y_j'\left(\frac{\pi_{ij}-\pi_i\pi_j}{\pi_i\pi_j}\right)$$

เมื่อ

$$y' = y_i - R_{HT}x_i$$

3. ค่าประมาณความแปรปรวนของตัวประมาณค่าเฉลี่ยอัตราส่วนเท่ากับ

$$v(\bar{y}_{rHT}) = \frac{1}{N^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{y_i'y_j'}{\pi_{ij}}\left(\frac{\pi_{ij}-\pi_i\pi_j}{\pi_i\pi_j}\right)$$

เมื่อ

$$y_i' = y_i - \hat{R}_{HT}x_i$$

พิสูจน์

1. กำหนดให้ $\hat{R}_{HT} = \frac{\sum_{i=1}^{n}\frac{y_i}{\pi_i}}{\sum_{i=1}^{n}\frac{x_i}{\pi_i}} = \frac{\hat{\tau}_y}{\hat{\tau}_x} = \frac{\bar{y}_{HT}}{\bar{x}_{HT}}$ (Cochran, 1977, p. 271) ซึ่งจะเห็นได้ว่า $\hat{R}_{HT}$ คือค่าเฉลี่ย $\bar{y}_{HT}$ หารด้วย $\bar{x}_{HT}$ หรือค่าประมาณผลรวม $\hat{\tau}_y$ หารด้วย $\hat{\tau}_x$ โดยที่ค่าประมาณการแจกแจงสิ่งตัวอย่างของ $\hat{R}_{HT}$ จะมีลักษณะเบ้ และค่อนข้างยุ่งยากมากกว่าค่าผลรวม $\hat{\tau}_y$ เนื่องจากทั้งตัวตั้ง $\hat{\tau}_y$ และตัวหาร $\hat{\tau}_x$ จะเปลี่ยนแปลงจากตัวอย่างชุดหนึ่งไปยังอีกตัวอย่างชุดหนึ่งโดยเฉพาะขนาดตัวอย่างน้อย ดังนั้นโดยปกติ $\hat{R}_{HT}$ จะเป็นตัวประมาณที่เอนเอียงสำหรับ R เนื่องจาก

$$\hat{\tau}_x = \tau_x + \left(\hat{\tau}_x - \tau_x\right) = \tau_x\left(1+\frac{\hat{\tau}_x-\tau_x}{\tau_x}\right) \text{ และ } \hat{\tau}_y = \tau_y + \left(\hat{\tau}_y - \tau_y\right) = \tau_y\left(1+\frac{\hat{\tau}_y-\tau_y}{\tau_y}\right)$$

ดังนั้น

$$\hat{R}_{HT} = \frac{\hat{\tau}_y}{\hat{\tau}_x}$$

$$= \frac{\tau_y}{\tau_x}\left(1+\frac{\hat{\tau}_y-\tau_y}{\tau_y}\right)\left(1+\frac{\hat{\tau}_x-\tau_x}{\tau_x}\right)^{-1}$$

$$= \frac{\tau_y}{\tau_x}\left(1+\frac{\hat{\tau}_y-\tau_y}{\tau_y}\right)\left(1-\frac{\hat{\tau}_x-\tau_x}{\tau_x}+\frac{(\hat{\tau}_x-\tau_x)^2}{\tau_x^2}+\dots\right)$$

$$= R\left(1-\frac{\hat{\tau}_x-\tau_x}{\tau_x}+\frac{\hat{\tau}_y-\tau_y}{\tau_y}+\frac{(\hat{\tau}_x-\tau_x)^2}{\tau_x^2}-\frac{\hat{\tau}_x-\tau_x}{\tau_x}\frac{\hat{\tau}_y-\tau_y}{\tau_y}+\dots\right)$$

เนื่องจาก $E(\hat{R}_{HT}) \neq R$ ดังนั้น $\hat{R}_{HT}$ เป็นตัวประมาณที่เอนเอียงของ R สำหรับตัวประมาณค่าเฉลี่ยแบบอัตราส่วน $\bar{y}_{rHT} = \hat{R}_{HT}\mu_x$ จะได้ว่า

$$E(\bar{y}_{rHT}) = E(\hat{R}_{HT}\mu_x)$$

$$= \mu_x E(\hat{R}_{HT})$$

$$\neq \mu$$

2. เนื่องจาก $\hat{R}_{HT}$ เป็นตัวประมาณที่เอนเอียงสำหรับ R ดังนั้น

$$MSE\left(\hat{R}_{HT}\right) = E\left(\hat{R}_{HT} - R\right)^2$$

$$= E\left(\frac{\hat{\tau}_y}{\hat{\tau}_x} - R\right)^2$$

$$= E\left(\frac{\hat{\tau}_y - R\hat{\tau}_x}{\hat{\tau}_x}\right)^2$$

เนื่องจาก $\hat{\tau}_y$ และ $\hat{\tau}_x$ ไม่ใช่ค่าคงที่ จะประมาณ $\hat{\tau}_x$ ด้วย τ_x และเนื่องจาก

$$E(\hat{\tau}_y) = E\left(\sum_{i=1}^{n}\frac{y_i}{\pi_i}\right) = E\left(\sum_{i=1}^{N}\frac{y_i l_i}{\pi_i}\right) = \sum_{i=1}^{N}\frac{y_i(\pi_i)}{\pi_i} = \sum_{i=1}^{N} y_i = \tau_y$$

$$E(\hat{\tau}_x) = E\left(\sum_{i=1}^{n}\frac{x_i}{\pi_i}\right) = E\left(\sum_{i=1}^{N}\frac{x_i l_i}{\pi_i}\right) = \sum_{i=1}^{N}\frac{x_i(\pi_i)}{\pi_i} = \sum_{i=1}^{N} x_i = \tau_x$$

และ

$$E(\hat{\tau}_y - R\hat{\tau}_x) = E\left(\hat{\tau}_y - \frac{\tau_y}{\tau_x}\hat{\tau}_x\right) = \left(\tau_y - \frac{\tau_y}{\tau_x}\tau_x\right) = 0$$

เพราะฉะนั้น

$$MSE\left(\hat{R}_{HT}\right) \cong \frac{1}{\tau_x^2} E\left(\hat{\tau}_y - R\hat{\tau}_x\right)^2$$

$$= \frac{1}{\tau_x^2} E\left[\left(\hat{\tau}_y - R\hat{\tau}_x\right) - E\left(\hat{\tau}_y - R\hat{\tau}_x\right)\right]^2$$

$$= \frac{1}{\tau_x^2} V\left(\hat{\tau}_y - R\hat{\tau}_x\right)$$

กำหนดให้ $y_i' = y_i - R_{HT}x_i$

จะได้ว่า $\hat{\tau}_y - R\hat{\tau}_x = \sum_{i=1}^{n}\frac{y_i}{\pi_i} - R\sum_{i=1}^{n}\frac{x_i}{\pi_i} = \sum_{i=1}^{n}\frac{y_i - Rx_i}{\pi_i} = \sum_{i=1}^{n}\frac{y_i'}{\pi_i} = \sum_{i=1}^{n} D_i$

โดยที่ $D_i = \dfrac{y_i'}{\pi_i}$

ดังนั้น

$$MSE\left(\hat{R}_{HT}\right) = \frac{1}{\tau_x^2} V\left(\sum_{i=1}^{n} D_i\right)$$

$$= \frac{1}{\tau_x^2} V\left(\sum_{i=1}^{N} D_i l_i\right)$$

$$= \frac{1}{\tau_x^2}\left(\sum_{i=1}^{N} Var[D_i l_i] + \sum_{i=1}^{N}\sum_{j\neq i}^{N} Cov[D_i l_i, D_j l_j]\right)$$

$$= \frac{1}{\tau_x^2}\left(\sum_{i=1}^{N} D_i^2 Var[l_i] + \sum_{i=1}^{N}\sum_{j\neq i}^{N} D_i D_j Cov[l_i l_j]\right)$$

$$=\frac{1}{\tau_x^2}\left(\sum_{i=1}^{N} D_i^2.\pi_i(1-\pi_i)+\sum_{i=1}^{N}\sum_{j\neq i}^{N} D_iD_j.(\pi_{ij}-\pi_i\pi_j)\right)$$

$$=\frac{1}{\tau_x^2}\left(\sum_{i=1}^{N}\sum_{i=j} D_iD_j.(\pi_{ij}-\pi_i\pi_j)+\sum_{i=1}^{N}\sum_{i\neq j}^{N} D_iD_j.(\pi_{ij}-\pi_i\pi_j)\right)$$

$$=\frac{1}{\tau_x^2}\left(\sum_{i=1}^{N}\sum_{j=1}^{N} D_iD_j.(\pi_{ij}-\pi_i\pi_j)\right)$$

$$=\frac{1}{\tau_x^2}\sum_{i=1}^{N}\sum_{j=1}^{N} y_i'y_j'\left(\frac{\pi_{ij}-\pi_i\pi_j}{\pi_i\pi_j}\right)$$

$$\cong V(\bar{y}_{rHT})$$

ดังนั้นความแปรปรวนของตัวประมาณค่าเฉลี่ยอัตราส่วนประมาณได้

$$V(\bar{y}_{rHT})=\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} y_i'y_j'\left(\frac{\pi_{ij}-\pi_i\pi_j}{\pi_i\pi_j}\right)$$

3. กำหนดให้

$$v(\bar{y}_{rHT})=\frac{1}{N^2}\sum_{i=1}^{n}\sum_{j=1}^{n} \frac{y_i'y_j'}{\pi_{ij}}\left(\frac{\pi_{ij}-\pi_i\pi_j}{\pi_i\pi_j}\right)$$

เนื่องจาก

$$E(v(\bar{y}_{rHT}))=E\left(\frac{1}{N^2}\left(\sum_{i=1}^{N}\sum_{j=1}^{N} \frac{y_i'y_j'l_il_j}{\pi_{ij}}\left(\frac{\pi_{ij}-\pi_i\pi_j}{\pi_i\pi_j}\right)\right)\right)$$

$$=\left(\frac{1}{N^2}\left(\sum_{i=1}^{N}\sum_{j=1}^{N} \frac{y_i'y_j'l_il_j}{\pi_{ij}}\left(\frac{\pi_{ij}-\pi_i\pi_j}{\pi_i\pi_j}\right)\right)\right);E(l_il_j)=\pi_{ij}$$

$$=\frac{1}{N^2}\left(\sum_{i=1}^{N}\sum_{j=1}^{N} \frac{y_i'y_j'.(\pi_{ij})}{\pi_{ij}}\left(\frac{\pi_{ij}-\pi_i\pi_j}{\pi_i\pi_j}\right)\right)$$

จะได้ว่า

$$E(v(\bar{y}_{rHT})) = \frac{1}{N^2}\left(\sum_{i=1}^{N}\sum_{j=1}^{N} y_i' y_j' \left(\frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j}\right)\right)$$

$$= V(\bar{y}_{rHT})$$

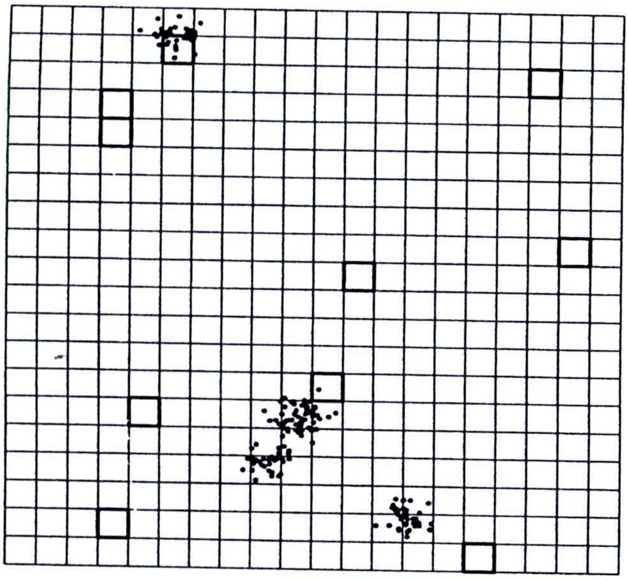
ในทางปฏิบัติเนื่องจากจะไม่ทราบ R ดังนั้นจะประมาณด้วย $\hat{R}$ และ y_i' จะประมาณด้วย $y_i - \hat{R}_{HT}x_i$

ทฤษฎีการสุ่มตัวอย่างแบบกลุ่มปรับเมื่อสุ่มหน่วยตัวอย่างอย่างง่าย

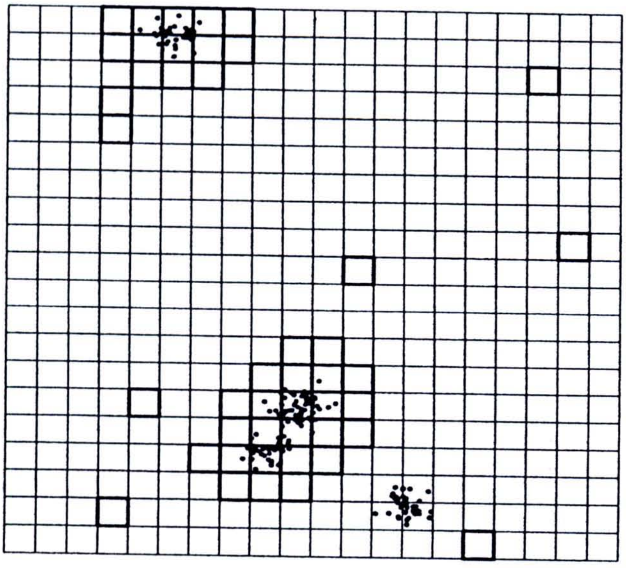
การเลือกหน่วยตัวอย่างในแผนแบบที่ยังไม่ปรับจะกระทำเสร็จก่อนการเก็บรวบรวมข้อมูลจากหน่วยตัวอย่างเหล่านั้น แต่สำหรับแผนการสุ่มตัวอย่างแบบกลุ่มปรับ มีหลักการว่า ลักษณะของข้อมูลที่เก็บรวบรวมได้ น่าจะมีส่วนช่วยทำให้ตัวประมาณมีคุณภาพดีขึ้นโดยการปรับกระบวนการเลือกตัวอย่างให้สามารถนำลักษณะข้อมูลที่เก็บรวบรวมได้มากำหนดหน่วยตัวอย่างที่เก็บรวบรวมเพิ่มขึ้น ทั้งในเชิงจำนวน และลักษณะของหน่วยที่จะเลือกเป็นตัวอย่างเพิ่มเติม เช่น กรณีของการประมาณจำนวนสัตว์ที่อาศัยอยู่ในพื้นที่หนึ่งในลักษณะที่ค่อนข้างเกาะกลุ่มกัน หากเลือกตัวอย่าง และพบสัตว์ประเภทนี้อยู่ในหน่วยตัวอย่างใด โอกาสที่จะพบในหน่วยใกล้เคียงหรือพื้นที่ข้างเคียงจะมีมากขึ้น ดังนั้นจึงควรเลือกหน่วยตัวอย่างในบริเวณใกล้เคียงเพิ่มเติม โดยหลักการ แผนการสุ่มตัวอย่างแบบกลุ่มปรับเป็นแผนแบบที่เลือกหน่วยตัวอย่างโดยกระบวนการสุ่มตัวอย่างด้วยความน่าจะเป็นแบบเท่ากัน เช่น การสุ่มตัวอย่างแบบง่าย การสุ่มตัวอย่างแบบมีระบบ และการสุ่มตัวอย่างแบบมีชั้นภูมิ เป็นต้น ซึ่งจะเลือกสุ่มตัวอย่างด้วยวิธีใดจะขึ้นอยู่กับหลายปัจจัย ไม่ว่าจะเป็นลักษณะของประชากรที่ศึกษา ขนาดตัวอย่าง งบประมาณ รวมทั้งระยะเวลา แต่สำหรับงานวิจัยนี้ใช้วิธีการสุ่มตัวอย่างแบบง่าย

แผนการสุ่มตัวอย่างแบบกลุ่มปรับเมื่อสุ่มหน่วยตัวอย่างอย่างง่าย คือแผนการสุ่มตัวอย่างที่ปรับกระบวนการสุ่มตัวอย่าง ให้สามารถนำลักษณะของข้อมูลที่เก็บได้มากำหนดหน่วยตัวอย่าง โดยการสุ่มหน่วยตัวอย่างจะใช้วิธีการสุ่มแบบง่าย เช่น กำหนดพื้นที่ที่ต้องการศึกษา และใช้วิธีการสุ่มแบบง่ายในการสำรวจสัตว์ หากพื้นที่ใดพบสัตว์เป็นไปตามเกณฑ์หรือเงื่อนไขที่กำหนด หน่วยทั้งหมดภายในส่วนที่ใกล้เคียงจะถูกเพิ่มเข้าเป็นตัวอย่าง และเป็นค่าสังเกตโดยจะขยายลักษณะเช่นนี้ไปเรื่อย ๆ จนกว่าจะไม่พบสิ่งที่สนใจหรือไม่เป็นไปตามเงื่อนไขที่กำหนด และบางครั้งการขยายการสุ่มหน่วยตัวอย่างในแต่ละกลุ่ม อาจทำให้เกิดการรวมเป็นเครือข่ายเดียวกัน ซึ่ง

ลักษณะการเก็บหน่วยตัวอย่างทั้งหมดที่ได้จากการสุ่มตัวอย่างภายใต้แผนแบบของการเลือกหน่วยตัวอย่างที่ประกอบด้วยการรวมส่วนที่ใกล้เคียงหลาย ๆ ส่วนจะเรียกว่า **กลุ่ม** โดยภายในกลุ่มเป็นการเก็บรวบรวมตัวอย่างย่อยของหน่วยเรียกว่า **เครือข่าย** สำหรับหน่วยตัวอย่างที่สุ่มได้แล้วไม่พบค่าสังเกตจะให้เครือข่ายนั้นมีขนาดเป็น 1 และถ้าหน่วยใดไม่เป็นไปตามเงื่อนไข แต่อยู่ในส่วนใกล้เคียงของหน่วยที่สอดคล้องตามเงื่อนไขจะเรียกหน่วยนั้นว่า **หน่วยขอบ** ความคิดพื้นฐานของแผนการสุ่มตัวอย่างแบบง่ายสามารถอธิบายได้ดังภาพที่ 2 ที่เป็นการประมาณค่าเฉลี่ยของจุดที่สุ่มโดยแบ่งบริเวณที่ศึกษาออกเป็น 400 พื้นที่ย่อย และทำการสุ่มหน่วยตัวอย่างด้วยแผนการสุ่มตัวอย่างแบบง่ายขนาด 10 หน่วย ซึ่งแสดงให้เห็นในภาพที่ 2(a) เมื่อสังเกตพบสิ่งที่สนใจจำนวนหนึ่งหรือมากกว่าในหน่วยที่ถูกเลือก หน่วยใกล้เคียงที่ถัดไปทางด้านบน ล่าง ขวา และซ้าย จะถูกเพิ่มเป็นตัวอย่าง เมื่อกระบวนการเสร็จสิ้นสมบูรณ์ ตัวอย่างจะประกอบด้วย 45 หน่วย แสดงได้ด้วยภาพ 2 (b) (Thompson, 1990, pp. 1050-1059)



(a)



(b)

ภาพ 2 แสดงแผนการสุ่มตัวอย่างแบบกลุ่มปรับ ในการประมาณค่าของจำนวนจุดสังเกต ในบริเวณที่ศึกษา 400 พื้นที่ย่อย

a) การสุ่มตัวอย่าง 10 หน่วย

b) การเพิ่มหน่วยที่ใกล้เคียงเนื่องจากสุ่มพบ 1 หรือมากกว่า 1 ค่าสังเกตของประชากรในหน่วยที่สุ่ม โดยจะทำการขยายหน่วยถัดไปทางด้านบน ล่าง ขวา และซ้าย แต่ถ้าไม่พบสิ่งที่สนใจก็จะไม่เพิ่มหน่วยถัดไป

แผนการสุ่มตัวอย่างแบบกลุ่มปรับ เป็นแผนแบบที่ใช้สำหรับงานด้านการสำรวจตัวอย่าง ซึ่งสิ่งที่สนใจในพื้นที่ที่จะทำการศึกษานั้นมีลักษณะเฉพาะ โดยโครงสร้างของประชากรค่อนข้างกระจุกตัว และเกาะกลุ่มกัน ดังนั้นตัวประมาณที่ใช้จึงต้องปรับให้มีความเหมาะสมกับแผนแบบ โดย ในปี ค.ศ. 1990 Thompson ได้เสนอตัวประมาณค่าเฉลี่ยที่ดัดแปลงมาจากตัวประมาณ Horvitz–Thompson และตัวประมาณค่าเฉลี่ยที่ดัดแปลงมาจากตัวประมาณ Hansen–Hurwitz

สำหรับงานวิจัยครั้งนี้ทำการศึกษาโดยใช้ตัวประมาณค่าเฉลี่ยที่ดัดแปลงมาจากตัวประมาณ Horvitz–Thompson ซึ่งเป็นตัวประมาณที่ใช้ความน่าจะเป็นไม่เท่ากันแบบไม่แทนที่ในรูปแบบของวิธีประมาณค่าเฉลี่ยแบบง่าย วิธีประมาณค่าเฉลี่ยแบบอัตราส่วน และวิธีประมาณค่าเฉลี่ยแบบอัตราส่วนที่ปรับปรุงโดยวิธีการของ Rao-Blackwell ซึ่งก่อนที่จะกล่าวถึงรายละเอียดและวิธีประมาณค่าดังกล่าวควรทำความเข้าใจเกี่ยวกับสัญลักษณ์ที่ใช้ในการประมาณค่าก่อนเพื่อไม่ให้เกิดความสับสนเกี่ยวกับการใช้สัญลักษณ์

โดยปกติแผนการสุ่มตัวอย่างแบบทั่วไปจะทำการสุ่มหน่วยตัวอย่างจากประชากร และจะนำค่าที่สนใจศึกษาที่ได้จากการสุ่มหน่วยตัวอย่างมาใช้ในการคำนวณค่า สัญลักษณ์ที่ใช้คือหน่วยตัวอย่างที่ i และค่าตัวแปรที่สนใจศึกษาคือ y_i ตามที่ได้กล่าวถึงในทฤษฎีแบบพื้นฐาน แต่สำหรับแผนการสุ่มตัวอย่างแบบกลุ่มปรับเมื่อทำการสุ่มหน่วยตัวอย่างจากประชากรแล้วจะทำการพิจารณาหน่วยที่อยู่บริเวณใกล้เคียงรวมเป็นหน่วยตัวอย่างด้วย ซึ่งการขยายลักษณะนี้ทำให้เกิดการรวมเป็นเครือข่าย ดังนั้นสัญลักษณ์ที่ใช้ในแผนการสุ่มตัวอย่างแบบกลุ่มปรับจึงใช้สัญลักษณ์เครือข่ายที่ k ที่มีสมาชิกหน่วยตัวอย่างที่ i ประกอบอยู่ในเครือข่ายดังกล่าว โดยมีรายละเอียดของตัวประมาณ และหลักการดังนี้

1. ตัวประมาณค่าเฉลี่ยแบบง่ายที่ดัดแปลงมาจากตัวประมาณ Horvitz–Thompson

Thompson (1990) ได้ศึกษาวิธีการประมาณค่าเฉลี่ย และความแปรปรวนของตัวประมาณค่าเฉลี่ยแบบง่ายภายใต้แผนการสุ่มตัวอย่างแบบกลุ่มปรับโดยมีรายละเอียดดังทฤษฎี 6

ทฤษฎี 6 การประมาณค่าเฉลี่ยแบบง่ายที่ดัดแปลงมาจากตัวประมาณ Horvitz–Thompson

1. ตัวประมาณที่ไม่เอนเอียงของค่าเฉลี่ยตัวอย่าง

$$\bar{y}^*_{HT} = \frac{1}{N}\sum_{k=1}^{\nu}\frac{u_{y_k}}{\alpha_k}$$

เมื่อ $\alpha_k = 1 - \binom{N-m_k}{n} \Big/ \binom{N}{n}$

โดยที่

α_k ความน่าจะเป็นที่หน่วยตัวอย่างที่ i อยู่ในเครือข่ายที่ k

u_{y_k} เป็นผลรวมของสิ่งที่สนใจ (y) ในเครือข่ายที่ k

N เป็นจำนวนของพื้นที่ย่อยทั้งหมด

n เป็นจำนวนพื้นที่ย่อยตัวอย่าง

m_k เป็นจำนวนของพื้นที่ย่อยในเครือข่ายที่ k

ν เป็นจำนวนเครือข่ายที่แตกต่างกันในตัวอย่าง

2. ความแปรปรวนของค่าประมาณค่าเฉลี่ยตัวอย่าง

$$V(\bar{y}^*_{HT}) = \frac{1}{N^2}\sum_{k=1}^{K}\sum_{h=1}^{K} u_{y_k} u_{y_h}\left(\frac{\alpha_{kh} - \alpha_k\alpha_h}{\alpha_k\alpha_h}\right)$$

เมื่อ $\alpha_{kh} = 1 - \left\{\binom{N-m_k}{n} + \binom{N-m_h}{n} - \binom{N-m_k-m_h}{n}\right\} \Big/ \binom{N}{n}$

โดยที่

α_h ความน่าจะเป็นที่หน่วยตัวอย่างที่ j อยู่ในเครือข่ายที่ h

α_{kh} เป็นความน่าจะเป็นร่วมที่หน่วยตัวอย่างที่ i และ j อยู่ในเครือข่ายที่ k และ h ตามลำดับ

u_{y_h} เป็นผลรวมของสิ่งที่สนใจ (y) ในเครือข่ายที่ h

3. ค่าประมาณความแปรปรวนของค่าประมาณค่าเฉลี่ยตัวอย่าง

$$v(\bar{y}^*_{HT}) = \frac{1}{N^2}\sum_{k=1}^{\nu}\sum_{h=1}^{\nu} \frac{u_{y_k} u_{y_h}}{\alpha_{kh}}\left(\frac{\alpha_{kh} - \alpha_k\alpha_h}{\alpha_k\alpha_h}\right)$$

พิสูจน์

1. จากแผนการสุ่มตัวอย่างแบบพื้นฐานด้วยความน่าจะเป็นไม่เท่ากันจะทำการสุ่มหน่วยตัวอย่าง n หน่วยจากประชากร N หน่วย และจะนำค่าที่สนใจศึกษา (y_i) ที่ได้จากการสุ่มหน่วยตัวอย่างมาใช้ในการคำนวณ โดยมีค่าความน่าจะเป็น π_i เข้ามาเกี่ยวข้อง แต่สำหรับแผนการสุ่มตัวอย่างแบบกลุ่มปรับจะทำการสุ่มพื้นที่ย่อยจากพื้นที่ทั้งหมดซึ่งภายในพื้นที่ย่อยอาจมี

หลายค่าสังเกตหรือหลายหน่วยตัวอย่าง ดังนั้นจึงปรับสัญลักษณ์ เป็นเครือข่ายที่ k และจะนำค่าที่ได้จากการสุ่มพื้นที่ย่อยที่เป็นไปตามเงื่อนไข (u_{y_k}) มาคำนวณโดยที่เครือข่ายที่นำมาคำนวณต้องไม่ซ้ำกัน (ν) และสัญลักษณ์ π_i จะเปลี่ยนเป็น α_k ดังนี้

$$\bar{y}^*_{HT} = \frac{1}{N}\sum_{i=1}^{n}\frac{y_i}{\pi_i}$$

$$= \frac{1}{N}\sum_{k=1}^{\nu}\frac{u_{y_k}}{\alpha_k}$$

$$= \frac{1}{N}\sum_{k=1}^{K}\frac{u_{y_k} I_k}{\alpha_k}$$

สำหรับ I_k มีคุณสมบัติการแจกแจงแบบเบอร์นูลีหรือ $I_k \sim Ber(\alpha_k)$ ดังนั้น $P(I_k = 1) = \alpha_k$, $P(I_k = 0) = 1 - \alpha_k$

และ

$$E(I_k) = 1.P(I_k = 1) + 0.P(I_k = 0) = \alpha_k$$

ในทำนองเดียวกันจะได้ $E(I_h) = \alpha_h$

เพราะฉะนั้น

$$E(\bar{y}^*_{HT}) = \frac{1}{N}\sum_{k=1}^{K}\frac{u_{y_k}}{\alpha_k}E(I_k)$$

$$= \frac{1}{N}\sum_{k=1}^{K}\frac{u_{y_k}}{\alpha_k}(\alpha_k)$$

$$= \frac{1}{N}\sum_{k=1}^{K}u_{y_k}$$

$$= \frac{1}{N}\sum_{i=1}^{N} y_i$$

$$= \mu$$

2. นอกจากนี้ I_k ยังมีคุณสมบัติดังนี้

$$V(I_k) = E(I_k^2) - [E(I_k)]^2$$

$$= E(I_k) - [E(I_k)]^2$$

$$= \alpha_k - \alpha_k^2$$

$$= \alpha_k(1 - \alpha_k)$$

$$E(I_k I_h) = \sum_{\text{all k,all h}} I_k I_h p(I_k I_h)$$

$$= 0(0)p(I_k = 0, I_h = 0) + 1(0)p(I_k = 1, I_h = 0) +$$

$$+0(1)p(I_k = 0, I_h = 0) + 1(1)p(I_k = 1, I_h = 1)$$

$$= p(I_k = 1, I_h = 1)$$

$$= \alpha_{kh}$$

และ

$$Cov(I_{k,} I_h) = E(I_k I_h) - E(I_k)E(I_h)$$

$$= \alpha_{kh} - \alpha_k \alpha_h$$

กำหนดให้ $Z_k = \frac{u_{y_k}}{\alpha_k}$ และ $Z_h = \frac{u_{y_h}}{\alpha_h}$ เพื่อใช้ในการพิสูจน์ $V(\bar{y}_{HT}^{*})$ จะได้ว่า

$$V(\bar{y}_{HT}^{*}) = V\left(\frac{1}{N}\sum_{k=1}^{K} Z_k I_k\right)$$

$$= \frac{1}{N^2}\left(\sum_{k=1}^{K} Var[Z_k I_k] + \sum_{k=1}^{K}\sum_{h\neq k}^{K} Cov[Z_k I_k, Z_h I_h]\right)$$

$$= \frac{1}{N^2}\left(\sum_{k=1}^{K} Z_k^2 Var[I_k] + \sum_{k=1}^{K}\sum_{h\neq k}^{K} Z_k Z_h Cov[I_k I_h]\right)$$

$$= \frac{1}{N^2}\left(\sum_{k=1}^{K} Z_k^2 .\alpha_k (1-\alpha_k) + \sum_{k=1}^{K}\sum_{h\neq k}^{K} Z_k Z_h .(\alpha_{kh} - \alpha_k \alpha_h)\right)$$

$$= \frac{1}{N^2}\left(\sum_{k=1}^{K}\sum_{h=k} Z_k Z_h .(\alpha_{kh} - \alpha_k \alpha_h) + \sum_{k=1}^{K}\sum_{h\neq k}^{K} Z_k Z_h .(\alpha_{kh} - \alpha_k \alpha_h)\right)$$

$$= \frac{1}{N^2}\left(\sum_{k=1}^{K}\sum_{h=1}^{K} Z_k Z_h .(\alpha_{kh} - \alpha_k \alpha_h)\right)$$

$$= \frac{1}{N^2}\sum_{k=1}^{K}\sum_{h=1}^{K} u_{y_k} u_{y_h}\left(\frac{\alpha_{kh} - \alpha_k \alpha_h}{\alpha_k \alpha_h}\right)$$

3. ค่าประมาณความแปรปรวนของ $v(\bar{y}_{HT}^{*})$ เป็นดังนี้

$$v(\bar{y}_{HT}^{*}) = \frac{1}{N^2}\sum_{k=1}^{v}\sum_{h=1}^{v} \frac{u_{y_k} u_{y_h}}{\alpha_{kh}}\left(\frac{\alpha_{kh} - \alpha_k \alpha_h}{\alpha_k \alpha_h}\right)$$

ดังนั้น

$$E(v(\bar{y}_{HT}^{*})) = E\left(\frac{1}{N^2}\left(\sum_{k=1}^{K}\sum_{h=1}^{K} \frac{u_{y_k} u_{y_h} .I_k I_h}{\alpha_{kh}}\left(\frac{\alpha_{kh} - \alpha_k \alpha_h}{\alpha_k \alpha_h}\right)\right)\right)$$

$$= \frac{1}{N^2}\left(\sum_{k=1}^{K}\sum_{h=1}^{K} \frac{u_{y_k} u_{y_h} .E(I_k I_h)}{\alpha_{kh}}\left(\frac{\alpha_{kh} - \alpha_k \alpha_h}{\alpha_k \alpha_h}\right)\right) ; E(I_k I_h) = \alpha_{kh}$$

$$= \frac{1}{N^2}\left(\sum_{k=1}^{K}\sum_{h=1}^{K}\frac{u_{y_k}u_{y_h}.(\alpha_{kh})}{\alpha_{kh}}\left(\frac{\alpha_{kh}-\alpha_k\alpha_h}{\alpha_k\alpha_h}\right)\right)$$

$$= \frac{1}{N^2}\left(\sum_{k=1}^{K}\sum_{h=1}^{K}u_{y_k}u_{y_h}\left(\frac{\alpha_{kh}-\alpha_k\alpha_h}{\alpha_k\alpha_h}\right)\right)$$

$$= V(\overline{y}^*_{HT})$$

4. ในการประมาณค่าความน่าจะเป็น α_k, α_h และ α_{kh} ที่ใช้สำหรับการประมาณค่าเฉลี่ย และความแปรปรวนสามารถกระทำได้ดังนี้

α_k คือ ความน่าจะเป็นที่หน่วยตัวอย่างที่ i อยู่ในเครือข่ายที่ k

ดังนั้นเราสามารถกำหนดให้ α_k (Thompson and Seber, 1996, p. 96) มีค่าเท่ากับ

$$\alpha_k = 1 - \binom{N-m_k}{n} \Big/ \binom{N}{n}$$

ในทำนองเดียวกัน

$$\alpha_h = 1 - \binom{N-m_h}{n} \Big/ \binom{N}{n}$$

ถ้ากำหนดให้ A และ B แทนเหตุการณ์ที่หน่วยตัวอย่างที่ i และ j อยู่ในเครือข่ายที่ k และ h ตามลำดับ ดังนั้น $P(A) = \alpha_k$ และ $P(B) = \alpha_h$

แต่เนื่องจาก $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

และ $P(A \cup B) = 1 - \binom{N-m_k-m_h}{n} \Big/ \binom{N}{n}$

จะได้ว่า $P(A \cap B) = P(A) + P(B) - P(A \cup B)$

$$= \alpha_k + \alpha_h - \left(1 - \binom{N-m_k-m_h}{n} \Big/ \binom{N}{n}\right)$$

$$= \alpha_k + \alpha_h + \binom{N-m_k-m_h}{n} \Big/ \binom{N}{n} - 1$$

$$= 1 - \frac{\binom{N-m_k}{n}}{\binom{N}{n}} + 1 - \frac{\binom{N-m_h}{n}}{\binom{N}{n}} + \frac{\binom{N-m_k-m_h}{n}}{\binom{N}{n}} - 1$$

$$= 1 - \frac{\binom{N-m_k}{n} + \binom{N-m_h}{n} - \binom{N-m_k-m_h}{n}}{\binom{N}{n}}$$

หรือ

$$\alpha_{kh} = P(A \cap B) = 1 - \left\{ \binom{N-m_k}{n} + \binom{N-m_h}{n} - \binom{N-m_k-m_h}{n} \right\} \Big/ \binom{N}{n}$$

2. **ตัวประมาณค่าเฉลี่ยแบบอัตราส่วนที่ดัดแปลงมาจากตัวประมาณ** Horvitz-Thompson

Dryver and Chao (2007) ได้ศึกษาวิธีการประมาณค่าเฉลี่ย และความแปรปรวนของตัวประมาณค่าเฉลี่ยแบบอัตราส่วนภายใต้แผนการสุ่มตัวอย่างแบบกลุ่มปรับโดยมีรายละเอียดดังทฤษฎี 7

ทฤษฎี 7 การประมาณค่าเฉลี่ยแบบอัตราส่วนที่ดัดแปลงมาจากตัวประมาณ Horvitz-Thompson (Chao, 2004)

1. ตัวประมาณค่าเฉลี่ยแบบอัตราส่วนที่เอนเอียงสำหรับค่าเฉลี่ยประชากร μ เท่ากับ

$$\bar{y}_{R.HT} = \frac{\bar{y}^*_{HT}}{\bar{x}^*_{HT}} \mu_x = \hat{R}_{ACS} \mu_x$$

เมื่อ

$$\bar{y}^*_{HT} = \frac{1}{N} \sum_{k=1}^{\nu} \frac{u_{y_k}}{\alpha_k} \text{ , } \bar{x}^*_{HT} = \frac{1}{N} \sum_{k=1}^{\nu} \frac{u_{x_k}}{\alpha_k} \text{ และ } \mu_x = \frac{1}{N} \sum_{i=1}^{N} x_i$$

2. ความแปรปรวนของตัวประมาณค่าเฉลี่ยแบบอัตราส่วนเท่ากับ

$$V(\overline{y}_{R.HT}) = \frac{1}{N^2}\sum_{k=1}^{K}\sum_{h=1}^{K} u'_{y_k} u'_{y_h} \left(\frac{\alpha_{kh} - \alpha_k \alpha_h}{\alpha_k \alpha_h} \right)$$

เมื่อ

$$u'_{y_k} = u_{y_k} - R_{ACS} u_{x_k}$$

3. ค่าประมาณความแปรปรวนของตัวประมาณค่าเฉลี่ยอัตราส่วนเท่ากับ

$$v(\overline{y}_{R.HT}) = \frac{1}{N^2}\sum_{k=1}^{\nu}\sum_{h=1}^{\nu} \frac{u'_{y_k} u'_{y_h}}{\alpha_{kh}} \left(\frac{\alpha_{kh} - \alpha_k \alpha_h}{\alpha_k \alpha_h} \right)$$

เมื่อ

$$u'_{y_k} = u_{y_k} - \hat{R}_{ACS} u_{x_k}$$

พิสูจน์

1. กำหนดให้ $\hat{R}_{ACS} = \dfrac{\sum_{k=1}^{\nu} \frac{u_{y_k}}{\alpha_k}}{\sum_{k=1}^{\nu} \frac{u_{x_k}}{\alpha_k}} = \dfrac{\hat{\tau}_y}{\hat{\tau}_x} = \dfrac{\overline{y}^*_{HT}}{\overline{x}^*_{HT}}$ ซึ่งจะเห็นได้ว่า $\hat{R}_{ACS}$ คือค่าเฉลี่ย $\overline{y}^*_{HT}$ หารด้วย $\overline{x}^*_{HT}$ หรือค่าประมาณผลรวม $\hat{\tau}_y$ หารด้วย $\hat{\tau}_x$ โดยที่การแจกแจงสิ่งตัวอย่างของ $\hat{R}_{ACS}$ จะมีลักษณะเบ้ และค่อนข้างยุ่งยากมากกว่าค่าผลรวม $\hat{\tau}_y$ เนื่องจากทั้งตัวตั้ง $\hat{\tau}_y$ และตัวหาร $\hat{\tau}_x$ จะเปลี่ยนแปลงจากตัวอย่างชุดหนึ่งไปยังอีกตัวอย่างชุดหนึ่งโดยเฉพาะขนาดตัวอย่างน้อย ดังนั้นโดยปกติ $\hat{R}_{ACS}$ จะเป็นตัวประมาณที่เอนเอียงสำหรับ R เนื่องจาก

$$\hat{\tau}_x = \tau_x + \left(\hat{\tau}_x - \tau_x\right) = \tau_x\left(1 + \frac{\hat{\tau}_x - \tau_x}{\tau_x}\right) \text{ และ } \hat{\tau}_y = \tau_y + \left(\hat{\tau}_y - \tau_y\right) = \tau_y\left(1 + \frac{\hat{\tau}_y - \tau_y}{\tau_y}\right)$$

ดังนั้น

$$\hat{R}_{ACS} = \frac{\hat{\tau}_y}{\hat{\tau}_x}$$

$$= \frac{\tau_y}{\tau_x}\left(1 + \frac{\hat{\tau}_y - \tau_y}{\tau_y}\right)\left(1 + \frac{\hat{\tau}_x - \tau_x}{\tau_x}\right)^{-1}$$

$$= \frac{\tau_y}{\tau_x}\left(1+\frac{\hat{\tau}_y-\tau_y}{\tau_y}\right)\left(1-\frac{\hat{\tau}_x-\tau_x}{\tau_x}+\frac{(\hat{\tau}_x-\tau_x)^2}{\tau_x^2}+\ldots\right)$$

$$= R\left(1-\frac{\hat{\tau}_x-\tau_x}{\tau_x}+\frac{\hat{\tau}_y-\tau_y}{\tau_y}+\frac{(\hat{\tau}_x-\tau_x)^2}{\tau_x^2}-\frac{\hat{\tau}_x-\tau_x}{\tau_x}\frac{\hat{\tau}_y-\tau_y}{\tau_y}+\ldots\right)$$

เนื่องจาก $E(\hat{R}_{ACS}) \neq R$ ดังนั้น $\hat{R}_{ACS}$ เป็นตัวประมาณที่เอนเอียงของ R สำหรับตัวประมาณค่าเฉลี่ยแบบอัตราส่วน $\bar{y}_{R.HT} = \hat{R}_{ACS}\mu_x$ จะได้ว่า

$$E(\bar{y}_{R.HT}) = E(\hat{R}_{ACS}\mu_x)$$

$$= \mu_x E(\hat{R}_{ACS})$$

$$\neq \mu$$

2. เนื่องจาก $\hat{R}_{ACS}$ เป็นตัวประมาณที่เอนเอียงสำหรับ R ดังนั้น

$$MSE\left(\hat{R}_{ACS}\right) = E\left(\hat{R}_{ACS} - R\right)^2$$

$$= E\left(\frac{\hat{\tau}_y}{\hat{\tau}_x} - R\right)^2$$

$$= E\left(\frac{\hat{\tau}_y - R\hat{\tau}_x}{\hat{\tau}_x}\right)^2$$

เนื่องจาก $\hat{\tau}_y$ และ $\hat{\tau}_x$ ไม่ใช่ค่าคงที่ จะประมาณ $\hat{\tau}_x$ ด้วย τ_x และเนื่องจาก

$$E(\hat{\tau}_y) = E\left(\sum_{k=1}^{\nu}\frac{u_{y_k}}{\alpha_k}\right) = E\left(\sum_{k=1}^{K}\frac{u_{y_k}I_k}{\alpha_k}\right) = \sum_{k=1}^{K}\frac{u_{y_k}(\alpha_k)}{\alpha_k} = \sum_{i=1}^{K}u_{y_k} = \tau_y$$

$$E(\hat{\tau}_x) = E\left(\sum_{k=1}^{\nu}\frac{u_{x_k}}{\alpha_k}\right) = E\left(\sum_{k=1}^{K}\frac{u_{x_k}I_k}{\alpha_k}\right) = \sum_{k=1}^{K}\frac{u_{x_k}(\alpha_k)}{\alpha_k} = \sum_{i=1}^{K}u_{x_k} = \tau_x$$

และ

$$E(\hat{\tau}_y - R\hat{\tau}_x) = E\left(\hat{\tau}_y - \frac{\tau_y}{\tau_x}\hat{\tau}_x\right) = \left(\tau_y - \frac{\tau_y}{\tau_x}\tau_x\right) = 0$$

เพราะฉะนั้น

$$MSE\left(\hat{R}_{ACS}\right) \cong \frac{1}{\tau_x^2} E\left(\hat{\tau}_y - R\hat{\tau}_x\right)^2$$

$$= \frac{1}{\tau_x^2} E\left[\left(\hat{\tau}_y - R\hat{\tau}_x\right) - E\left(\hat{\tau}_y - R\hat{\tau}_x\right)\right]^2$$

$$= \frac{1}{\tau_x^2} V\left(\hat{\tau}_y - R\hat{\tau}_x\right)$$

กำหนดให้

$$u'_{y_k} = u_{y_k} - R_{ACS} u_{x_k}$$

จะได้ว่า

$$\hat{\tau}_y - R\hat{\tau}_x = \sum_{k=1}^{\nu} \frac{u_{y_k}}{\alpha_k} - R\sum_{k=1}^{\nu} \frac{u_{x_k}}{\alpha_k} = \sum_{k=1}^{\nu} \frac{u_{y_k} - Ru_{x_k}}{\alpha_k} = \sum_{k=1}^{\nu} \frac{u'_{y_k}}{\alpha_k} = \sum_{k=1}^{\nu} D_k$$

โดยที่

$$D_k = \frac{u'_{y_k}}{\alpha_k}$$

ดังนั้น

$$MSE\left(\hat{R}_{ACS}\right) = \frac{1}{\tau_x^2} V\left(\sum_{k=1}^{\nu} D_k\right)$$

$$= \frac{1}{\tau_x^2} V\left(\sum_{k=1}^{K} D_k I_k\right)$$

$$=\frac{1}{\tau_x^2}\left(\sum_{k=1}^{K}Var[D_kI_k]+\sum_{k=1}^{K}\sum_{h\neq k}^{K}Cov[D_kI_k,D_hI_h]\right)$$

$$=\frac{1}{\tau_x^2}\left(\sum_{k=1}^{K}D_k^2Var[I_k]+\sum_{k=1}^{K}\sum_{h\neq k}^{K}D_kD_hCov[I_kI_h]\right)$$

$$=\frac{1}{\tau_x^2}\left(\sum_{k=1}^{K}D_k^2.\alpha_k(1-\alpha_k)+\sum_{k=1}^{K}\sum_{h\neq k}^{K}D_kD_h.(\alpha_{kh}-\alpha_k\alpha_h)\right)$$

$$=\frac{1}{\tau_x^2}\left(\sum_{k=1}^{K}\sum_{h=k}^{K}D_kD_h.(\alpha_{kh}-\alpha_k\alpha_h)+\sum_{k=1}^{K}\sum_{k\neq h}^{K}D_kD_h.(\alpha_{kh}-\alpha_k\alpha_h)\right)$$

$$=\frac{1}{\tau_x^2}\left(\sum_{k=1}^{K}\sum_{h=k}^{K}D_kD_h.(\alpha_{kh}-\alpha_k\alpha_h)\right)$$

$$=\frac{1}{\tau_x^2}\sum_{k=1}^{K}\sum_{h=1}^{K}u'_{y_k}u'_{y_h}\left(\frac{\alpha_{kh}-\alpha_k\alpha_h}{\alpha_k\alpha_h}\right)$$

$$\cong V(\bar{y}_{R.HT})$$

ดังนั้นความแปรปรวนของตัวประมาณค่าเฉลี่ยอัตราส่วนประมาณได้

$$V(\bar{y}_{R.HT})=\frac{1}{N^2}\sum_{k=1}^{K}\sum_{h=1}^{K}u'_{y_k}u'_{y_h}\left(\frac{\alpha_{kh}-\alpha_k\alpha_h}{\alpha_k\alpha_h}\right)$$

3. ค่าประมาณความแปรปรวนของ $v(\bar{y}_{R.HT})$ เป็นดังนี้

$$v(\bar{y}_{R.HT})=\frac{1}{N^2}\sum_{k=1}^{v}\sum_{h=1}^{v}\frac{u'_{y_k}u'_{y_h}}{\alpha_{kh}}\left(\frac{\alpha_{kh}-\alpha_k\alpha_h}{\alpha_k\alpha_h}\right)$$

ดังนั้น

$$E(v(\bar{y}_{R.HT}))=E\left(\frac{1}{N^2}\left(\sum_{k=1}^{K}\sum_{h=1}^{K}\frac{u'_{y_k}u'_{y_h}.I_kI_h}{\alpha_{kh}}\left(\frac{\alpha_{kh}-\alpha_k\alpha_h}{\alpha_k\alpha_h}\right)\right)\right)$$

$$= \frac{1}{N^2}\left(\sum_{k=1}^{K}\sum_{h=1}^{K}\frac{u'_{y_k}u'_{y_h}.E(I_k I_h)}{\alpha_{kh}}\left(\frac{\alpha_{kh}-\alpha_k\alpha_h}{\alpha_k\alpha_h}\right)\right) ; E(I_k I_h)=\alpha_{kh}$$

$$= \frac{1}{N^2}\left(\sum_{k=1}^{K}\sum_{h=1}^{K}\frac{u'_{y_k}u'_{y_h}.(\alpha_{kh})}{\alpha_{kh}}\left(\frac{\alpha_{kh}-\alpha_k\alpha_h}{\alpha_k\alpha_h}\right)\right)$$

จะได้ว่า

$$E(v(\bar{y}_{R.HT})) = \frac{1}{N^2}\left(\sum_{k=1}^{K}\sum_{h=1}^{K}u'_{y_k}u'_{y_h}\left(\frac{\alpha_{kh}-\alpha_k\alpha_h}{\alpha_k\alpha_h}\right)\right)$$

$$= V(\bar{y}_{R.HT})$$

ในทางปฏิบัติเนื่องจากจะไม่ทราบ R ดังนั้นจะประมาณด้วย $\hat{R}$ และ u'_{y_k} จะประมาณด้วย $u_{y_k} - \hat{R}_{ACS}u_{x_k}$

จากวิธีประมาณค่าแบบอัตราส่วน จะเห็นได้ว่าเป็นวิธีประมาณค่าที่นำค่าสังเกตที่ได้จากการสุ่มพื้นที่ย่อยแล้วขยายเครือข่ายมาคำนวณค่า โดยมีการใช้ประโยชน์จากตัวแปรช่วย แต่ถ้าในกรณีที่ค่าสังเกตหรือสิ่งที่สนใจศึกษาไม่เป็นไปตามเงื่อนไขที่กำหนด ก็จะตกเป็นหน่วยขอบ ซึ่งถ้าใช้วิธีประมาณค่าแบบอัตราส่วนก็อาจจะได้สาระไม่ครบถ้วนมากนัก ดังนั้นเพื่อแก้ไขปัญหาดังกล่าวจึงมีผู้ที่ศึกษาวิธีประมาณค่า โดยนำหน่วยขอบมาร่วมในการพิจารณา ซึ่งวิธีประมาณค่าดังกล่าวได้มีการใช้ทฤษฎี Rao-Blackwell และก่อนที่จะกล่าวถึงวิธีประมาณค่าแบบอัตราส่วนที่ปรับปรุงโดยวิธีการ Rao-Blackwell ควรทำความเข้าใจเกี่ยวกับตัวแปรที่ใช้ในการประมาณค่าของทั้ง 2 วิธี สำหรับวิธีประมาณค่าแบบอัตราส่วนที่ปรับปรุงโดยวิธีการ Rao-Blackwell จะมีการนำค่าหน่วยขอบมาพิจารณาคำนวณค่า ซึ่งทำให้เกิดเงื่อนไขในการคำนวณ โดยถ้าทำการสุ่มพื้นที่ย่อยแล้วพบหน่วยขอบจะใช้สูตรในการคำนวณที่แตกต่างไปจากวิธีการประมาณค่าแบบอัตราส่วน แต่ถ้าสุ่มพื้นที่ย่อยแล้วไม่พบหน่วยขอบคือเป็นกรณีอื่น ๆ สูตรในการประมาณค่าก็จะเป็นวิธีการเดียวกับการประมาณค่าแบบอัตราส่วนนั่นเอง การกำหนดเงื่อนไขในลักษณะนี้จะทำให้ได้สาระครบถ้วนมากกว่าเดิม เนื่องจากมีการใช้ข้อมูลทั้งหน่วยขอบ และเครือข่ายอย่างครบถ้วนซึ่งต่างจากวิธีประมาณค่าแบบอัตราส่วนที่ใช้เพียงเครือข่ายเท่านั้น โดยหน่วยขอบดังกล่าวอาจจะมีสารประโยชน์ต่อการประมาณค่า ถ้ามองข้ามข้อมูลดังกล่าวอาจทำให้สูญเสียข้อมูลที่สำคัญ ส่งผลให้การประมาณค่ามีประสิทธิภาพลดลง ซึ่งการประมาณค่าที่มีการใช้ข้อมูลที่ครบถ้วนนี้จะทำให้

เกิดเป็นสถิติพอเพียง โดยสถิติที่พอเพียงบอกให้เราทราบว่าสารสนเทศเกี่ยวกับพารามิเตอร์มีอยู่ครบถ้วนแล้ว ไม่จำเป็นต้องไปค้นหาวิธีประมาณที่มีคุณสมบัติเหมาะสมจากที่อื่น (ประชุม สุวัตถี, 2545, หน้า 210) ซึ่งรายละเอียดของตัวประมาณ และหลักการมีดังนี้

3. **ตัวประมาณค่าเฉลี่ยแบบอัตราส่วนที่ปรับปรุงโดยวิธีการ Rao-Blackwell**

Chao, Lin and Chiang (2008) ได้ทำการศึกษาวิธีการประมาณค่าเฉลี่ย และความแปรปรวนของตัวประมาณค่าเฉลี่ยแบบอัตราส่วนที่ปรับปรุงโดยวิธีการ Rao-Blackwell ซึ่งมีรายละเอียดดังทฤษฎี 8

ทฤษฎี 8 การประมาณค่าเฉลี่ยแบบอัตราส่วนที่ปรับปรุงโดยวิธีการ Rao-Blackwell

1. ตัวประมาณค่าเฉลี่ยแบบอัตราส่วนที่ปรับปรุงโดยวิธีการ Rao-Blackwell

$$\bar{y}^{+}_{R.HT} = \frac{\bar{y}^{+}_{HT}}{\bar{x}^{+}_{HT}}\mu_x = \frac{\sum_{k=1}^{K}\frac{u_{y_k}}{\alpha_k}\left(1-\sum_{i\in\psi_k} e_i\right)+\frac{n}{N}\bar{y}_e}{\sum_{k=1}^{K}\frac{u_{x_k}}{\alpha_k}\left(1-\sum_{i\in\psi_k} e_i\right)+\frac{n}{N}\bar{x}_e}\mu_x = \hat{R}^{+}_{ACS}\mu_x$$

เมื่อ

$$\bar{y}^{+}_{HT} = E\left(\bar{y}^{*}_{HT} \mid D^{+} = d^{+}\right) = \frac{1}{N}\sum_{k=1}^{v}\frac{m_k . w^{+}_{y_k}}{\alpha_k}$$

และ

$$\bar{x}^{+}_{HT} = E\left(\bar{x}^{*}_{HT} \mid D^{+} = d^{+}\right) = \frac{1}{N}\sum_{k=1}^{v}\frac{m_k . w^{+}_{x_k}}{\alpha_k}$$

กำหนดให้ d^{+} เป็นสถิติพอเพียงของ μ (Dryver and Thompsom , 2005, p.159) มีค่าเท่ากับ

$$d^{+} = \left\{\left(i, y_i, f_i\right), \left(j, y_j\right); i \in s_c, j \in s_{\bar{c}}\right\}$$

โดยที่ D^{+} แทนสเปซตัวอย่างของ d^{+}

f_i เป็นจำนวนครั้งที่เครือข่ายที่ i รวมอยู่ในตัวอย่างที่ได้จากการสุ่มตัวอย่างเริ่มต้น

s_c เป็นเซตของหน่วยขอบตัวอย่างที่เป็นไปตามเงื่อนไขที่กำหนด

$s_{\bar{c}}$ เป็นเซตของหน่วยขอบตัวอย่างที่ไม่เป็นไปตามเงื่อนไขที่กำหนด

$$w_{y_k}^{+} = \begin{cases} \bar{y}_e = \dfrac{\sum_{i \in s} e_i y_i}{e_s} & , if \sum_{i \in \psi_k} e_i = 1 \\ w_{y_k} & , if \sum_{i \in \psi_k} e_i = 0 \end{cases} \quad \text{และ} \quad w_{y_k} = \frac{1}{m_k} \sum u_{y_k}$$

กำหนดให้ $e_i = \begin{cases} 1, \text{ไม่พบตามเงื่อนไขแต่อยู่บริเวณใกล้เคียงกับส่วนที่เป็นไปตามเงื่อนไข} \\ 0, \text{อื่น ๆ} \end{cases}$

โดยที่

$\bar{y}_e$ เป็นค่าเฉลี่ยของค่า y สำหรับหน่วยขอบตัวอย่าง

e_s เป็นจำนวนหน่วยขอบตัวอย่างใน s

w_{y_k} เป็นค่าเฉลี่ยของค่า y ในเครือข่ายที่ k

2. ความแปรปรวนของตัวประมาณค่าเฉลี่ยแบบอัตราส่วนที่ปรับปรุงโดยวิธีการ Rao-Blackwell

$$V(\bar{y}_{R.HT}^{+}) = V(\bar{y}_{R.HT}) - \frac{1}{n^2} \sum_{d^+ \in D^+} P(d^+) \times \left(\frac{\left(e_{s_0}\right)}{\left(e_s\right)} \sum_{i \in s_0, e_i = 1} (y_i - R_{ACS} x_i)^2 \right.$$

$$\left. + \frac{e_{s_0}\left(e_{s_0} - 1\right)}{e_s\left(e_s - 1\right)} \sum_{i \in s_0, e_i = 1} \sum_{j \neq i} (y_i - R_{ACS} x_i)(y_j - R_{ACS} x_j) - e_{s_0}^2 (\bar{y}_e - R_{ACS} \bar{x}_e)^2 \right)$$

หรือ

$V(\bar{y}_{R.HT}^{+}) = V(\bar{y}_{R.HT}) - (\mu_{R.HT}^{+} - \mu_{R.HT})^2$ (Dryver and Chao, 2007)

โดยที่

$P(d^+)$ คือความน่าจะเป็นที่ $D^+ = d^+$

e_{s_0} เป็น จำนวนของหน่วยขอบตัวอย่างสำหรับการสุ่มตัวอย่างเริ่มต้น s_0

$\overline{x}_e$ เป็นค่าเฉลี่ยของค่า x สำหรับหน่วยขอบตัวอย่าง

3. ค่าประมาณความแปรปรวนของตัวประมาณค่าเฉลี่ยแบบอัตราส่วนที่ปรับปรุงโดยวิธีการ Rao-Blackwell

$$v(\overline{y}^{+}_{R.HT}) = v(\overline{y}_{R.HT}) - \frac{1}{n^2}\sum_{d^{+}\in D^{+}} P(d^{+}) \times \left(\frac{\left(e_{s_0}\right)}{\left(e_s\right)} \sum_{i\in s_0, e_i=1} (y_i - \hat{R}_{ACS}x_i)^2 \right.$$

$$\left. + \frac{e_{s_0}\left(e_{s_0}-1\right)}{e_s\left(e_s-1\right)} \sum_{i\in s_0, e_i=1}\sum_{j\neq i}(y_i - \hat{R}_{ACS}x_i)(y_j - \hat{R}_{ACS}x_j) - e_{s_0}^2(\overline{y}_e - \hat{R}_{ACS}\overline{x}_e)^2 \right)$$

หรือ

$v(\overline{y}^{+}_{R.HT}) = v(\overline{y}_{R.HT}) - (\overline{y}^{+}_{R.HT} - \overline{y}_{R.HT})^2$ (Dryver and Chao, 2007)

พิสูจน์

1. การพิสูจน์จากตัวประมาณค่าเฉลี่ยแบบอัตราส่วนเป็นตัวประมาณค่าเฉลี่ยแบบอัตราส่วนที่ปรับปรุงโดยวิธีการ Rao-Blackwell สามารถกระทำได้ดังนี้

$$\overline{y}^{+}_{R.HT} = E\left(\overline{y}_{R.HT} \mid D^{+} = d^{+}\right)$$

$$= \sum_{s_0\in S} \overline{y}^{+}_{R.HT}(s_0)P(S_0 = s_0 \mid D^{+} = d^{+})$$

โดยความน่าจะเป็นแบบมีเงื่อนไข $P(S_0 = s_0 \mid D^{+} = d^{+})$ สามารถเขียนให้อยู่ในรูปของ

$$\frac{1}{L}\sum_{s_0\in S} I\{g(s_0) = d^{+}\}$$

เมื่อ $I\{.\}$ เป็นตัวแปรบ่งชี้ และ $L = \sum_{s_0 \in S} I\{g(s_0) = d^+\}$ เป็นจำนวนผลรวมของการเลือกหมู่ (Combination) ที่เกี่ยวข้องกับ d^+ และการรวมของการเลือกหมู่ L จากหน่วยตัวอย่างใด ๆ หนึ่งหน่วยมีค่าเท่ากับ $I\left\{g(s_0) = d^+\right\} = \dfrac{\binom{e_s - 1}{e_{s_0} - 1}}{\binom{e_s}{e_{s_0}}} L$

ดังนั้น

$$\bar{y}^+_{R.HT} = \frac{1}{L} \sum_{s_0 \in S} I\{g(s_0) = d^+\} \bar{y}^+_{R.HT}(s_0)$$

$$= \frac{\binom{e_s - 1}{e_{s_0} - 1}}{\binom{e_s}{e_{s_0}}} \sum_{s_0 \in S} \bar{y}^+_{R.HT}(s_0)$$

$$= \frac{e_{s_0}}{e_s} \sum_{i \in s, e_i = 1} \frac{\sum_{k=1}^{K} \frac{u_{y_k}}{\alpha_k} \left(1 - \sum_{i \in \psi_k} e_i\right) + \frac{n}{N} e_i y_i}{\sum_{k=1}^{K} \frac{u_{x_k}}{\alpha_k} \left(1 - \sum_{i \in \psi_k} e_i\right) + \frac{n}{N} e_i x_i} \mu_x \quad ; \bar{y}_e = \frac{\sum_{i \in s} e_i y_i}{e_s} \text{ และ } \bar{x}_e = \frac{\sum_{i \in s} e_i x_i}{e_{s_0}}$$

$$= \frac{\sum_{k=1}^{K} \frac{u_{y_k}}{\alpha_k} \left(1 - \sum_{i \in \psi_k} e_i\right) + \frac{n}{N} \bar{y}_e}{\sum_{k=1}^{K} \frac{u_{x_k}}{\alpha_k} \left(1 - \sum_{i \in \psi_k} e_i\right) + \frac{n}{N} \bar{x}_e} \mu_x$$

และการหาตัวประมาณของ $\hat{R}^+_{ACS}$ สามารถกระทำได้ดังนี้

กำหนดให้ $\hat{R}^{+}_{ACS} = \dfrac{\sum_{k=1}^{v} \dfrac{m_k . w^{+}_{y_k}}{\alpha_k}}{\sum_{k=1}^{v} \dfrac{m_k . w^{+}_{x_k}}{\alpha_k}} = \dfrac{\hat{\tau}_y}{\hat{\tau}_x} = \dfrac{\overline{y}^{+}_{HT}}{\overline{x}^{+}_{HT}}$ ซึ่งจะเห็นได้ว่า $\hat{R}^{+}_{ACS}$ คือค่าเฉลี่ย $\overline{y}^{+}_{HT}$ หารด้วย $\overline{x}^{+}_{HT}$ หรือค่าประมาณผลรวม $\hat{\tau}_y$ หารด้วย $\hat{\tau}_x$ โดยที่การแจกแจงสิ่งตัวอย่างของ $\hat{R}^{+}_{ACS}$ จะมีลักษณะเบ้ และค่อนข้างยุ่งยากมากกว่าค่าผลรวม $\hat{\tau}_y$ เนื่องจากทั้งตัวตั้ง $\hat{\tau}_y$ และตัวหาร $\hat{\tau}_x$ จะเปลี่ยนแปลงจากตัวอย่างชุดหนึ่งไปยังอีกตัวอย่างชุดหนึ่งโดยเฉพาะขนาดตัวอย่างน้อย ดังนั้นโดยปกติ $\hat{R}^{+}_{ACS}$ จะเป็นตัวประมาณที่เอนเอียงสำหรับ R เนื่องจาก

$\hat{\tau}_x = \tau_x + \left(\hat{\tau}_x - \tau_x\right) = \tau_x \left(1 + \dfrac{\hat{\tau}_x - \tau_x}{\tau_x}\right)$ และ $\hat{\tau}_y = \tau_y + \left(\hat{\tau}_y - \tau_y\right) = \tau_y \left(1 + \dfrac{\hat{\tau}_y - \tau_y}{\tau_y}\right)$

ดังนั้น

$$\hat{R}^{+}_{ACS} = \frac{\tau_y}{\tau_x}\left(1 + \frac{\hat{\tau}_y - \tau_y}{\tau_y}\right)\left(1 + \frac{\hat{\tau}_x - \tau_x}{\tau_x}\right)^{-1}$$

$$= \frac{\tau_y}{\tau_x}\left(1 + \frac{\hat{\tau}_y - \tau_y}{\tau_y}\right)\left(1 - \frac{\hat{\tau}_x - \tau_x}{\tau_x} + \frac{(\hat{\tau}_x - \tau_x)^2}{\tau_x^2} + \ldots\right)$$

$$= R\left(1 - \frac{\hat{\tau}_x - \tau_x}{\tau_x} + \frac{\hat{\tau}_y - \tau_y}{\tau_y} + \frac{(\hat{\tau}_x - \tau_x)^2}{\tau_x^2} - \frac{\hat{\tau}_x - \tau_x}{\tau_x}\frac{\hat{\tau}_y - \tau_y}{\tau_y} + \ldots\right)$$

เนื่องจาก $E(\hat{R}^{+}_{ACS}) \neq R$ ดังนั้น $\hat{R}^{+}_{ACS}$ เป็นตัวประมาณที่เอนเอียงของ R

สำหรับตัวประมาณค่าเฉลี่ยแบบอัตราส่วน $\overline{y}^{+}_{R.HT} = \hat{R}^{+}_{ACS}\mu_x$ จะได้ว่า

$$E(\overline{y}^{+}_{R.HT}) = E(\hat{R}^{+}_{ACS}\mu_x)$$

$$= \mu_x E(\hat{R}^{+}_{ACS})$$

$$\neq \mu$$

2. เนื่องจาก $\bar{y}^{+}_{R.HT} = \dfrac{\bar{y}^{+}_{HT}}{\bar{x}^{+}_{HT}}\mu_x = \hat{R}^{+}_{ACS}\mu_x = \bar{y}^{+}_{HT} + \hat{R}_{ACS}(\mu_x - \bar{x}^{+}_{HT})$ (Thompson 2002, p. 77)

ประมาณค่า $\hat{R}$ ด้วย R จะได้

$$\bar{y}^{+}_{R.HT} \approx \bar{y}^{+}_{HT} + R_{ACS}(\mu_x - \bar{x}^{+}_{HT})$$

ดังนั้น

$$\bar{y}^{+}_{R.HT} - \mu \approx \bar{y}^{+}_{HT} + R_{ACS}(\mu_x - \bar{x}^{+}_{HT}) - \mu$$

$$= \bar{y}^{+}_{HT} + R_{ACS}\mu_x - R_{ACS}\bar{x}^{+}_{HT} - \mu$$

$$= \bar{y}^{+}_{HT} + \mu - R_{ACS}\bar{x}^{+}_{HT} - \mu \quad ; R_{ACS}\mu_x = \mu$$

$$= \bar{y}^{+}_{HT} - R_{ACS}\ \bar{x}^{+}_{HT}$$

เพราะฉะนั้น

$$V\ (\bar{y}^{+}_{R.HT} - \mu) = V(\bar{y}^{+}_{HT} - R_{ACS}\ \bar{x}^{+}_{HT})$$

$$V(\bar{y}^{+}_{R.HT}) = V(\bar{y}^{+}_{HT} - R_{ACS}\ \bar{x}^{+}_{HT})$$

ในทำนองเดียวกันจะได้

$$V(\bar{y}_{R.HT}) = V(\bar{y}^{*}_{HT} - R_{ACS}\ \bar{x}^{*}_{HT})$$

เนื่องจาก $\bar{y}^{+}_{R.HT} = \bar{y}^{+}_{HT} - R_{ACS}\bar{x}^{+}_{HT}$ และ $\bar{y}^{+}_{R.HT} = E\left(\bar{y}_{R.HT} \mid D^{+} = d^{+}\right)$
ดังนั้น

$$V(\bar{y}^{+}_{R.HT}) = V(E(\bar{y}^{*}_{HT} - R_{ACS}\bar{x}^{*}_{HT}) \mid D^{+}))$$

$$= V\left(\bar{y}^{*}_{HT} - R_{ACS}\bar{x}^{*}_{HT}\right) - E\left(V\left(\bar{y}^{*}_{HT} - R_{ACS}\bar{x}^{*}_{HT}\right) \mid D^{+}\right)$$

$$= V\left(\bar{y}_{R.HT}\right) - E\left(V\left(\bar{y}^{*}_{HT} - R_{ACS}\bar{x}^{*}_{HT}\right) \mid D^{+}\right)$$

$$= V(\bar{y}_{R.HT}) - E\left((\bar{y}^{*}_{HT} - R_{ACS}\bar{x}^{*}_{HT}) - (\bar{y}^{+}_{HT} - R_{ACS}\bar{x}^{+}_{HT})\right)^2$$

$$= V(\bar{y}_{R.HT}) - \frac{1}{n^2}\sum_{d^+ \in D^+}\frac{P(d^+)}{L(d^+)}\sum_{s_0 \in S} I\{g(s_0) = d^+\} \times$$
$$\left(\sum_{i \in s_0, e_i = 1}(y_i - R_{ACS}x_i) - e_{s_0}(\bar{y}_e - R_{ACS}\bar{x}_e)\right)^2$$

$$= V(\bar{y}_{R.HT}) - \frac{1}{n^2}\sum_{d^+ \in D^+}\frac{P(d^+)}{L(d^+)}\sum_{s_0 \in S} I\{g(s_0) = d^+\} \times$$
$$\left(\left\{\sum_{i \in s_0, e_i = 1}(y_i - R_{ACS}x_i)\right\}^2 - e_{s_0}^2(\bar{y}_e - R_{ACS}\bar{x}_e)^2\right)$$

$$= V(\bar{y}_{R.HT}) - \frac{1}{n^2}\sum_{d^+ \in D^+}\frac{P(d^+)}{L(d^+)}\sum_{s_0 \in S} I\{g(s_0) = d^+\} \times$$
$$\left(\sum_{i \in s_0, e_i = 1}(y_i - R_{ACS}x_i)^2 + \sum_{i \in s_0, e_i = 1}\sum_{j \neq i}(y_i - R_{ACS}x_i)(y_j - R_{ACS}x_j) - e_{s_0}^2(\bar{y}_e - R_{ACS}\bar{x}_e)^2\right)$$

$$= V(\bar{y}_{R.HT}) - \frac{1}{n^2}\sum_{d^+ \in D^+} P(d^+) \times \left(\frac{\binom{e_s - 1}{e_{s_0} - 1}}{\binom{e_s}{e_{s_0}}}\sum_{i \in s_0, e_i = 1}(y_i - R_{ACS}x_i)^2\right.$$
$$\left. + \frac{\binom{e_s - 2}{e_{s_0} - 2}}{\binom{e_s}{e_{s_0}}}\sum_{i \in s_0, e_i = 1}\sum_{j \neq i}(y_i - R_{ACS}x_i)(y_j - R_{ACS}x_j) - e_{s_0}^2(\bar{y}_e - R_{ACS}\bar{x}_e)^2\right)$$

$$= V(\bar{y}_{R.HT}) - \frac{1}{n^2}\sum_{d^+ \in D^+} P(d^+) \times \left(\frac{(e_{s_0})}{(e_s)}\sum_{i \in s_0, e_i = 1}(y_i - R_{ACS}x_i)^2\right.$$
$$\left. + \frac{e_{s_0}(e_{s_0} - 1)}{e_s(e_s - 1)}\sum_{i \in s_0, e_i = 1}\sum_{j \neq i}(y_i - R_{ACS}x_i)(y_j - R_{ACS}x_j) - e_{s_0}^2(\bar{y}_e - R_{ACS}\bar{x}_e)^2\right)$$

3. ค่าประมาณความแปรปรวนของ $v(\overline{y}^{+}_{R.HT})$ เป็นดังนี้

$$v(\overline{y}^{+}_{R.HT}) = v(\overline{y}_{R.HT}) - \frac{1}{n^2}\sum_{d^+\in D^+} P(d^+)\times\left(\frac{\left(e_{s_0}\right)}{\left(e_s\right)}\sum_{i\in s_0,e_i=1}(y_i-\hat{R}_{ACS}x_i)^2\right.$$

$$\left.+\frac{e_{s_0}\left(e_{s_0}-1\right)}{e_s\left(e_s-1\right)}\sum_{i\in s_0,e_i=1}\sum_{j\neq i}(y_i-\hat{R}_{ACS}x_i)(y_j-\hat{R}_{ACS}x_j)-e_{s_0}^2(\overline{y}_e-\hat{R}_{ACS}\overline{x}_e)^2\right)$$

เนื่องจาก

$$E\left(v(\overline{y}_{R.HT})\right) = V(\overline{y}_{R.HT})$$

ดังนั้น

$$V(\overline{y}^{+}_{R.HT}) = V(\overline{y}_{R.HT}) - \frac{1}{n^2}\sum_{d^+\in D^+} P(d^+)\times\left(\frac{\left(e_{s_0}\right)}{\left(e_s\right)}\sum_{i\in s_0,e_i=1}(y_i-R_{ACS}x_i)^2\right.$$

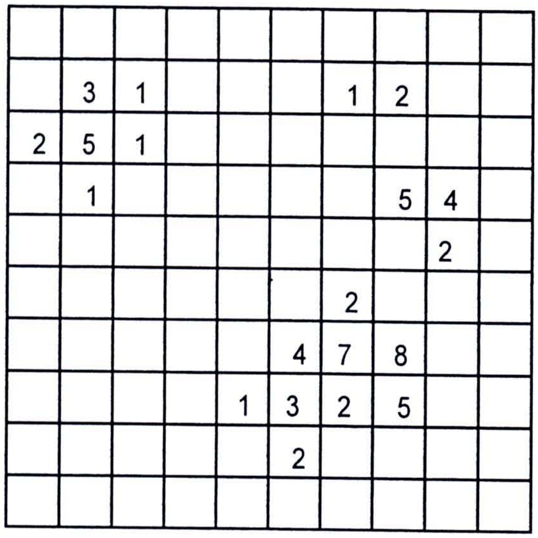
$$\left.+\frac{e_{s_0}\left(e_{s_0}-1\right)}{e_s\left(e_s-1\right)}\sum_{i\in s_0,e_i=1}\sum_{j\neq i}(y_i-R_{ACS}x_i)(y_j-R_{ACS}x_j)-e_{s_0}^2(\overline{y}_e-R_{ACS}\overline{x}_e)^2\right)$$

ในทางปฏิบัติเนื่องจากจะไม่ทราบ R_{ACS} ดังนั้นจะประมาณด้วย $\hat{R}_{ACS}$

4. ตัวอย่างการคำนวณ

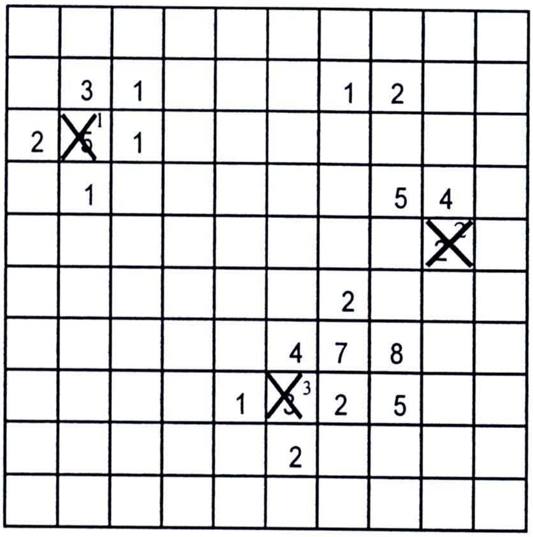
4.1 กรณีหน่วยขอบแยกกัน

ตัวอย่างที่ 1 ต้องการประมาณจำนวนประชากรกวางผาในป่าแห่งหนึ่งทางตอนเหนือของประเทศไทย โดยพบว่ากวางผามักจะอาศัยอยู่บริเวณที่เป็นเนินเขาหรือผาที่เต็มไปด้วยหิน ดังนั้นจึงกำหนดพื้นที่ที่จะทำการศึกษาที่เป็นเนินเขาหรือผาออกเป็น 100 ตร.หน่วย ซึ่งในป่าดังกล่าวมีจำนวนกวางผาแสดงดังภาพ 3



ภาพ 3 แสดงที่ตั้ง และจำนวนของประชากรกวางผากรณีหน่วยขอบแยกกัน

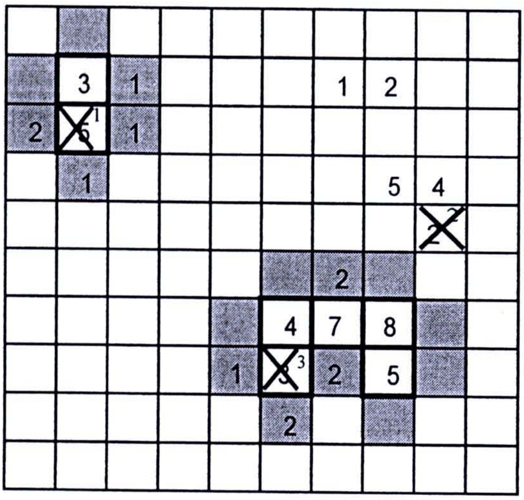
จากนั้นทำการสุ่มตัวอย่างด้วยแผนการสุ่มตัวอย่างแบบง่าย ขนาด 3 พื้นที่ย่อย $(n = 3)$ โดยสุ่มพื้นที่ได้ตรงตำแหน่งดังภาพ 4



ภาพ 4 แสดงตำแหน่งของการสุ่มพื้นที่ย่อยขนาด 3 หน่วยด้วยแผนการสุ่มตัวอย่างแบบง่าย

จากภาพ 4 สุ่มตัวอย่างครั้งที่ 1 ได้ตำแหน่งที่ 22 ครั้งที่ 2 ได้ตำแหน่งที่ 49 และครั้งที่ 3 ได้ตำแหน่งที่ 76

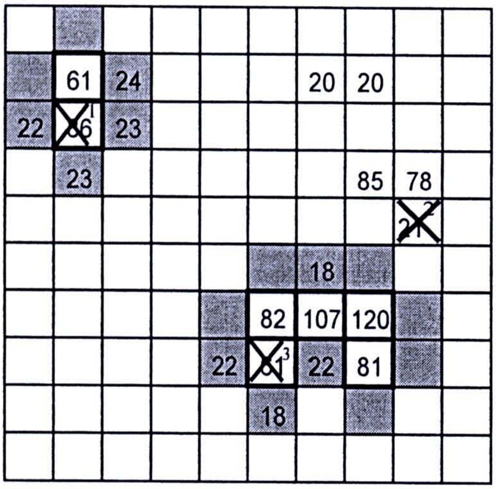
จากนั้นเมื่อทำการสุ่มพื้นที่ย่อยแล้วก็จะพิจารณาพื้นที่ย่อยบริเวณใกล้เคียง โดยที่เงื่อนไขของการรวมเป็นเครือข่ายคือ $y_i \geq 3$ โดยถ้าเป็นไปตามเงื่อนไขก็จะถูกรวมเป็นเครือข่าย แต่ถ้าไม่เป็นไปตามเงื่อนไขก็จะไม่พิจารณา และส่วนที่อยู่รอบเครือข่ายจะเรียกว่าหน่วยขอบ สามารถอธิบายได้ดังภาพ 5



ภาพ 5 แสดงการขยายพื้นที่ในลักษณะ ขึ้น-ลง-ซ้าย-ขวา จนกระทั่งไม่เป็นไปตามเงื่อนไข

จากภาพ 5 พบว่ามีพื้นที่ 3 หน่วยที่พบกวางผาได้แก่พื้นที่ที่ 1, 2 และ3 แต่เมื่อทำการพิจารณาตามเงื่อนไขถ้าพบกวางผามากกว่าหรือเท่ากับ 3 ตัวจะทำการขยายพื้นที่ในลักษณะ ขึ้น-ลง-ซ้าย และขวา ขยายไปเรื่อย ๆ จนกว่าจะไม่พบกวางผาตามเงื่อนไข ดังนั้นพื้นที่ที่ 2 จึงไม่เป็นไปตามเงื่อนไขที่กำหนด เพราะฉะนั้นจึงมี 2 เครือข่ายที่ได้จากการขยายตามเงื่อนไข

และจากการศึกษาพบว่าประชากรกวางผาจะพบมากบริเวณป่าที่มีความอุดมสมบูรณ์ และมีพุ่มไม้หนาทึบ ดังนั้นตัวแปรช่วยที่ใช้ในการศึกษาครั้งนี้คือจำนวนต้นไม้ที่ขึ้นหนาทึบบริเวณเนินเขาหรือหน้าผา แสดงได้ดังภาพ 6



ภาพ 6 แสดงจำนวนต้นไม้ที่ขึ้นหนาทึบบริเวณเนินเขาหรือหน้าผาที่กวางผาอาศัยอยู่

จากภาพ 5 และ 6 สามารถแทนในสูตรได้ดังนี้ โดยกำหนดให้

$N = 100$, $n = 3$, $v = 3$ (เนื่องจากไม่มีเครือข่ายที่ซ้ำกัน) $m_1 = 2$, $m_2 = 1$, $m_3 = 5$

ตัวแปร Y

ผลรวมของเครือข่ายที่ 1 (u_{y_1}) คือ $(3+5) = 8$ ผลรวมของเครือข่ายที่ 2 (u_{y_2}) คือ 2 ผลรวมของเครือข่ายที่ $3(u_{y_3})$ คือ $(4+7+8+3+5) = 27$

ผลรวมของหน่วยขอบเครือข่ายที่ 1 (y_{e_1}) คือ $(2+1+1+1+0+0) = 5$ ผลรวมของหน่วยขอบเครือข่ายที่ 2 (y_{e_2}) คือ $(0+0+0+0+0+2) = 2$ ผลรวมของหน่วยขอบเครือข่ายที่ 3 (y_{e_3}) คือ $(2+1+0+0+2+0+0+0+0+2) = 7$

ตัวแปร X

ผลรวมของเครือข่ายที่ 1 (u_{x_1}) คือ $(61+86) = 147$ ผลรวมของเครือข่ายที่ 2 (u_{x_2}) คือ 21 ผลรวมของเครือข่ายที่ 3 (u_{x_3}) คือ $(82+107+120+61+81) = 451$

ผลรวมของหน่วยขอบเครือข่ายที่ $1(x_{e_1})$ คือ $(24+23+23+22+0+0) = 92$ ผลรวมของหน่วยขอบเครือข่ายที่ 2 (x_{e_2}) คือ $(0+0+0+0+0+21) = 21$ ผลรวมของหน่วยขอบเครือข่ายที่ 3 (x_{e_3}) คือ $(18+22+0+0+18+0+0+0+0+22) = 80$

วิธีการประมาณค่าแบบง่าย

การประมาณจำนวนประชากรกวางผาโดยเฉลี่ยด้วยวิธีการประมาณค่าอย่างง่ายมีดังนี้

คำนวณค่าความน่าจะเป็นของเครือข่ายที่ได้จากการพิจารณาตามเงื่อนไขโดยที่ α_k มีค่าเท่ากับ

$$\alpha_k = 1 - \binom{N - m_k}{n} \Big/ \binom{N}{n}$$

$$\alpha_1 = 1 - \binom{100-2}{3} \Big/ \binom{100}{3} = 0.0594$$

$$\alpha_2 = 1 - \binom{100-1}{3} \Big/ \binom{100}{3} = 0.03$$

$$\alpha_3 = 1 - \binom{100-5}{3} \Big/ \binom{100}{3} = 0.144$$

คำนวณค่าความน่าจะเป็นร่วมของเครือข่ายที่ได้จากการพิจารณาตามเงื่อนไขโดยที่ α_{kh} มีค่าเท่ากับ

$$\alpha_{kh} = 1 - \left\{ \binom{N - m_k}{n} + \binom{N - m_h}{n} - \binom{N - m_k - m_h}{n} \right\} \Big/ \binom{N}{n}$$

$$\alpha_{12} = 1 - \left\{ \binom{100-2}{3} + \binom{100-1}{3} - \binom{100-2-1}{3} \right\} \Big/ \binom{100}{3} = 0.0012$$

$$\alpha_{13} = 1 - \left\{ \binom{100-2}{3} + \binom{100-5}{3} - \binom{100-2-5}{3} \right\} \Big/ \binom{100}{3} = 0.0059$$

$$\alpha_{23} = 1 - \left\{ \binom{100-1}{3} + \binom{100-5}{3} - \binom{100-1-5}{3} \right\} \Big/ \binom{100}{3} = 0.003$$

ดังนั้น

$$\bar{y}^{*}_{HT}=\frac{1}{N}\sum_{k=1}^{\nu}\frac{u_{y_k}}{\alpha_k}=\frac{1}{100}\sum_{k=1}^{3}\frac{u_{y_k}}{\alpha_k}=\frac{1}{100}\left[\left(\frac{8}{0.0594}\right)+\left(\frac{2}{0.03}\right)+\left(\frac{27}{0.144}\right)\right]=3.89$$

และค่าประมาณความแปรปรวนของการประมาณจำนวนประชากรกวางผาโดยเฉลี่ยมีค่าเท่ากับ

$$v(\bar{y}^{*}_{HT})=\frac{1}{N^2}\left[\sum_{k=1}^{\nu}\frac{(u_{y_k})^2}{\alpha_k}\left(\frac{1}{\alpha_k}-1\right)+2\sum_{k=1}^{\nu-1}\sum_{l=k+1}^{\nu}\frac{u_{y_k}u_{y_h}}{\alpha_{kh}}\left(\frac{\alpha_{kh}}{\alpha_k\alpha_h}-1\right)\right]$$

$$=\frac{1}{100^2}\left[\begin{aligned}&\left(\frac{(8)^2}{0.0594}\left(\frac{1}{0.0594}-1\right)+\frac{(2)^2}{0.03}\left(\frac{1}{0.03}-1\right)+\frac{(27)^2}{0.144}\left(\frac{1}{0.144}-1\right)\right)\\&+2\left(\frac{(8)(2)}{0.0012}\right)\left(\frac{0.0012}{(0.0594)(0.03)}-1\right)+2\left(\frac{(8)(27)}{0.0059}\right)\left(\frac{0.0059}{(0.0594)(0.144)}-1\right)\\&+2\left(\frac{(2)(27)}{0.003}\right)\left(\frac{0.003}{(0.03)(0.144)}-1\right)\end{aligned}\right]$$

$$=0.8992$$

วิธีการประมาณค่าแบบอัตราส่วน

การประมาณจำนวนประชากรกวางผาโดยเฉลี่ยด้วยวิธีการประมาณค่าแบบอัตราส่วนมีค่าเท่ากับ

$$\bar{y}_{R.HT}=\frac{\bar{y}^{*}_{HT}}{\bar{x}^{*}_{HT}}\mu_x \text{ โดยที่ } \bar{y}^{*}_{HT}=\frac{1}{N}\sum_{k=1}^{\upsilon}\frac{u_{y_k}}{\alpha_k},\ \bar{x}^{*}_{HT}=\frac{1}{N}\sum_{k=1}^{\upsilon}\frac{u_{x_k}}{\alpha_k},\ \mu_x=\frac{1}{N}\sum_{i=1}^{N}x_i$$

และ

$$\bar{y}^{*}_{HT}=\frac{1}{N}\sum_{k=1}^{\nu}\frac{u_{y_k}}{\alpha_k}=\frac{1}{100}\sum_{k=1}^{3}\frac{u_{y_k}}{\alpha_k}=\frac{1}{100}\left[\left(\frac{8}{0.0594}\right)+\left(\frac{2}{0.03}\right)+\left(\frac{27}{0.144}\right)\right]=3.89$$

$$\bar{x}^{*}_{HT}=\frac{1}{N}\sum_{k=1}^{\nu}\frac{u_{x_k}}{\alpha_k}=\frac{1}{100}\sum_{k=1}^{3}\frac{u_{x_k}}{\alpha_k}=\frac{1}{100}\left[\left(\frac{147}{0.0594}\right)+\left(\frac{21}{0.03}\right)+\left(\frac{451}{0.144}\right)\right]=63.067$$

$$\mu_x = \frac{1}{100}\sum_{i=1}^{100}(0+0+....+0) = 9.94$$

ดังนั้น

$$\bar{y}_{R.HT} = \frac{\bar{y}_{HT}^*}{\bar{x}_{HT}^*}\mu_x = \frac{3.89}{63.067}9.94 = 0.6131$$

และค่าประมาณความแปรปรวนของการประมาณจำนวนประชากรกวางผาโดยเฉลี่ยมีค่าเท่ากับ

$$v(\bar{y}_{R.HT}) = \frac{1}{N^2}\left[\sum_{k=1}^{\nu}\frac{\left(u'_{y_k}\right)^2}{\alpha_k}\left(\frac{1}{\alpha_k}-1\right) + 2\sum_{k=1}^{\nu-1}\sum_{l=k+1}^{\nu}\frac{u'_{y_k}u'_{y_h}}{\alpha_{kh}}\left(\frac{\alpha_{kh}}{\alpha_k\alpha_h}-1\right)\right]$$

เมื่อ

$u'_{y_k} = u_{y_k} - \hat{R}_{ACS}u_{x_k}$ โดยที่ $\hat{R}_{ACS} = \frac{\bar{y}_{HT}^*}{\bar{x}_{HT}^*}$

และ

$$\hat{R}_{ACS} = \frac{3.89}{63.067} = 0.0617$$

$$u'_{y_1} = 8 - (0.0617 \times 147) = -1.0699$$

$$u'_{y_2} = 2 - (0.0617 \times 21) = 0.7043$$

$$u'_{y_3} = 27 - (0.0617 \times 451) = -0.8267$$

ดังนั้น

$$v(\bar{y}_{R.HT}) = \frac{1}{100^2}\left[\begin{array}{l} \left(\frac{(-1.0699)^2}{0.0594}\left(\frac{1}{0.0594}-1\right)+\frac{(0.7043)^2}{0.03}\left(\frac{1}{0.03}-1\right)+\frac{(-0.8267)^2}{0.144}\left(\frac{1}{0.144}-1\right)\right) \\ +2\left(\frac{(-1.0699)(0.7043)}{0.0012}\right)\left(\frac{0.0012}{(0.0594)(0.03)}-1\right)+2\left(\frac{(-1.0699)(-0.8267)}{0.0059}\right)\left(\frac{0.0059}{(0.00594)(0.144)}-1\right) \\ +2\left(\frac{(0.7043)(-0.8267)}{0.003}\right)\left(\frac{0.003}{(0.03)(0.144)}-1\right) \end{array}\right]$$

$$= 0.1304$$

วิธีการประมาณค่าแบบอัตราส่วนที่ปรับปรุงโดยวิธีการ Rao-Blackwell

การประมาณจำนวนประชากรกวางผาโดยเฉลี่ยด้วยวิธีการประมาณค่าแบบอัตราส่วนมีค่าเท่ากับ

$$\bar{y}^{+}_{R.HT} = \frac{\bar{y}^{+}_{HT}}{\bar{x}^{+}_{HT}}\mu_x$$

โดยที่

$$\bar{y}^{+}_{HT} = \frac{1}{N}\sum_{k=1}^{\upsilon}\frac{m_k.w^{+}_{y_k}}{\alpha_k} \ , \ \bar{x}^{+}_{HT} = \frac{1}{N}\sum_{k=1}^{\upsilon}\frac{m_k.w^{+}_{x_k}}{\alpha_k} \ , \ \mu_x = \frac{1}{N}\sum_{i=1}^{N}x_i$$

เมื่อ

$$w^{+}_{y_k} = \begin{cases} w_{y_k} \\ \bar{y}_e = \dfrac{\sum_{i\in s} y_i}{e_s} \end{cases} , \ w_{y_k} = \frac{1}{m_k}\sum y_i$$

และ

$$\bar{y}^{+}_{HT} = \frac{1}{N}\sum_{k=1}^{\nu}\frac{m_k.w^{+}_{y_k}}{\alpha_k} = \frac{1}{100}\sum_{k=1}^{3}\frac{m_k.w^{+}_{y_k}}{\alpha_k} = \frac{1}{100}\left[\left(\frac{2\times\frac{8}{2}}{0.0594}\right)+\left(\frac{1\times\frac{2}{6}}{0.03}\right)+\left(\frac{5\times\frac{27}{5}}{0.144}\right)\right] = 3.3333$$

$$\overline{x}_{HT}^{+} = \frac{1}{N}\sum_{k=1}^{\nu}\frac{m_k.w_{x_k}^{+}}{\alpha_k} = \frac{1}{100}\sum_{k=1}^{3}\frac{m_k.w_{x_k}^{+}}{\alpha_k} = \frac{1}{100}\left[\left(\frac{2\times\frac{147}{2}}{0.0594}\right)+\left(\frac{1\times\frac{21}{6}}{0.03}\right)+\left(\frac{5\times\frac{451}{5}}{0.144}\right)\right] = 57.2336$$

$$\mu_x = \frac{1}{100}\sum_{i=1}^{100}(0+0+....+0) = 9.94$$

ดังนั้น

$$\overline{y}_{R.HT}^{+} = \frac{3.333}{57.2336}9.94 = 0.5788$$

และค่าประมาณความแปรปรวนของการประมาณจำนวนประชากรกวางผาโดยเฉลี่ย

$$v(\overline{y}_{R.HT}^{+}) \quad = v(\overline{y}_{R.HT}) - (\overline{y}_{R.HT}^{+} - \overline{y}_{R.HT})^2$$

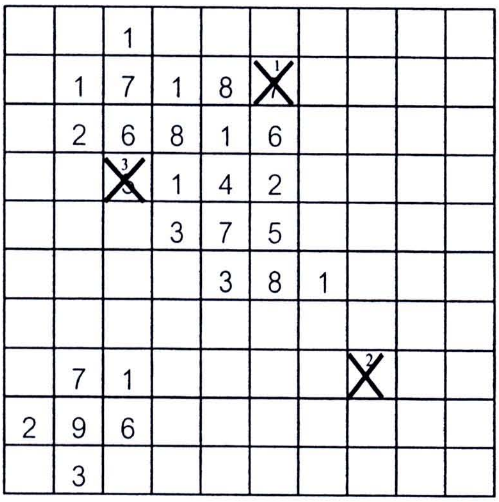
$$= 0.13037 - (0.5788 - 0.6131)^2$$
$$= 0.1292$$

4.2 กรณีหน่วยขอบร่วมกัน

ตัวอย่างที่ 2 ใช้แหล่งข้อมูลชุดเดียวกับตัวอย่างที่ 1 แต่ประชากรเป็นกรณีที่หน่วยขอบร่วมกัน

		1							
	1	7	1	8	7				
	2	6	8	1	6				
		5	1	4	2				
			3	7	5				
				3	8	1			
	7	1							
2	9	6							
	3								

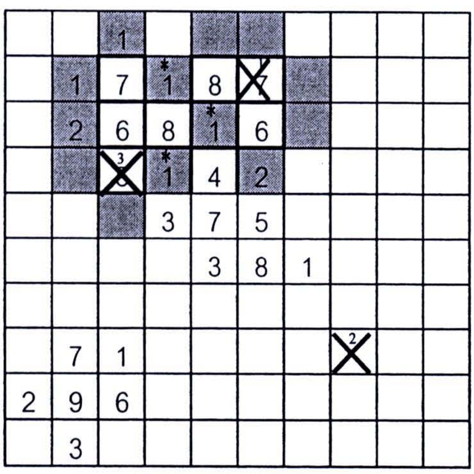
ภาพ 7 แสดงที่ตั้ง และจำนวนของประชากรกวางผากรณีหน่วยขอบร่วมกัน



ภาพ 8 แสดงตำแหน่งของการสุ่มตัวอย่างขนาด 3 หน่วยด้วยแผนการสุ่มตัวอย่างแบบง่าย

จากภาพ 8 สุ่มตัวอย่างครั้งที่ 1 ได้ตำแหน่งที่ 16 ครั้งที่ 2 ได้ตำแหน่งที่ 33 และครั้งที่ 3 ได้ตำแหน่งที่ 78

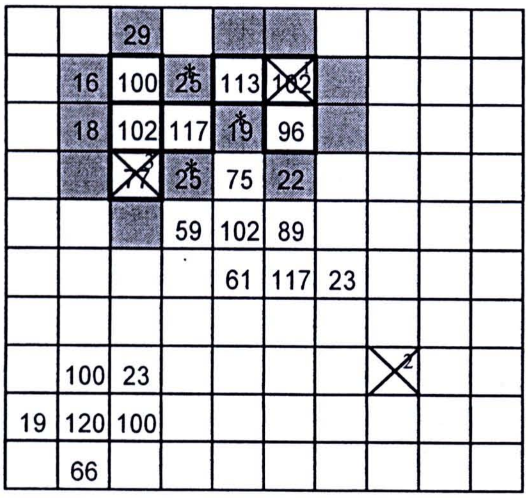
จากนั้นเมื่อทำการสุ่มพื้นที่ย่อยแล้วก็จะพิจารณาพื้นที่ย่อยบริเวณใกล้เคียง โดยที่เงื่อนไขของการรวมเป็นเครือข่ายคือ $y_i \geq 3$ โดยถ้าเป็นไปตามเงื่อนไขก็จะถูกรวมเป็นเครือข่าย แต่ถ้าไม่เป็นไปตามเงื่อนไขก็จะไม่พิจารณา และส่วนที่อยู่รอบเครือข่ายจะเรียกว่าหน่วยขอบ โดยที่ * คือส่วนที่มีหน่วยขอบร่วมกันระหว่างกลุ่ม สามารถจะอธิบายได้ดังภาพ 9



ภาพ 9 แสดงการขยายพื้นที่ในลักษณะ ขึ้น-ลง-ซ้าย-ขวา จนกระทั่งไม่เป็นไปตามเงื่อนไข

จากภาพ 9 พบว่ามีพื้นที่ 2 หน่วยที่พบกวางผาได้แก่พื้นที่ที่ 1 และ3 ส่วนพื้นที่ที่ 2 ไม่พบกวางผา และเมื่อทำการพิจารณาตามเงื่อนไขถ้าพบกวางผามากกว่าหรือเท่ากับ 3 ตัวจะทำการขยายพื้นที่ในลักษณะ ขึ้น-ลง-ซ้าย และขวา ขยายไปเรื่อย ๆ จนกว่าจะไม่พบกวางผาตามเงื่อนไข ดังนั้นจึงมีพื้นที่ 1 และ3 ที่เป็นไปตามเงื่อนไขที่กำหนด เพราะฉะนั้นจึงมี 2 เครือข่ายที่ได้จากการขยายตามเงื่อนไข

และจากการศึกษาพบว่าประชากรกวางผาจะพบมากบริเวณป่าที่มีความอุดมสมบูรณ์ และมีพุ่มไม้หนาทึบ ดังนั้นตัวแปรช่วยที่ใช้ในการศึกษาครั้งนี้คือจำนวนต้นไม้ที่ขึ้นหนาทึบบริเวณเนินเขาหรือหน้าผา แสดงได้ดังภาพ 10



ภาพ 10 แสดงจำนวนต้นไม้ที่ขึ้นหนาทึบบริเวณเนินเขาหรือหน้าผาที่กวางผาอาศัยอยู่

จากภาพ 9 และ 10 สามารถแทนในสูตรได้ดังนี้ โดยกำหนดให้

$N = 100$, $n = 3$, $\nu = 3$ (เนื่องจากไม่มีเครือข่ายที่ซ้ำกัน) $m_1 = 3$, $m_2 = 4$, $m_3 = 1$

ตัวแปร Y

ผลรวมของเครือข่ายที่ 1 (u_{y_1}) คือ $(8+7+6) = 21$ ผลรวมของเครือข่ายที่ 2 (u_{y_2}) คือ 0 ผลรวมของเครือข่ายที่ 3 (u_{y_3}) คือ $(7+6+8+5) = 26$

ผลรวมของหน่วยขอบเครือข่ายที่ 1 (y_{e_1}) คือ $(2+1+1+0+0+0+0) = 5$ ผลรวมของหน่วยขอบเครือข่ายที่ 2 (y_{e_2}) คือ 0 ผลรวมของหน่วยขอบเครือข่ายที่ 3 (y_{e_3}) คือ $(2+1+1+1+1+1+0+0) = 7$

ตัวแปร X

ผลรวมของเครือข่ายที่ 1 (u_{x_1}) คือ $(113+102+96)=311$ ผลรวมของเครือข่ายที่ 2 (u_{x_2}) คือ 0 ผลรวมของเครือข่ายที่ 3 (u_{x_3}) คือ $(100+102+117+77)=396$

ผลรวมของหน่วยขอบเครือข่ายที่ 1 (x_{e_1}) คือ $(25+19+22+0+0+0+0)=66$ ผลรวมของหน่วยขอบเครือข่ายที่ 2 (x_{e_2}) คือ 0 ผลรวมของหน่วยขอบเครือข่ายที่ 3 (x_{e_3}) คือ $(18+16+29+25+19+25+0+0)=132$

วิธีการประมาณค่าแบบง่าย

การประมาณจำนวนประชากรกวางผาโดยเฉลี่ยด้วยวิธีการประมาณค่าอย่างง่ายมีดังนี้

คำนวณค่าความน่าจะเป็นของเครือข่ายที่ได้จากการพิจารณาตามเงื่อนไขโดยที่ α_k มีค่าเท่ากับ

$$\alpha_k = 1-\binom{N-m_k}{n}\bigg/\binom{N}{n}$$

$$\alpha_1 = 1-\binom{100-3}{3}\bigg/\binom{100}{3} = 0.0882$$

$$\alpha_2 = 1-\binom{100-1}{3}\bigg/\binom{100}{3} = 0.0300$$

$$\alpha_3 = 1-\binom{100-4}{3}\bigg/\binom{100}{3} = 0.1164$$

คำนวณค่าความน่าจะเป็นร่วมของเครือข่ายที่ได้จากการพิจารณาตามเงื่อนไขโดยที่ α_{kh} มีค่าเท่ากับ

$$\alpha_{kh} = 1-\left\{\binom{N-m_k}{n}+\binom{N-m_h}{n}-\binom{N-m_k-m_h}{n}\right\}\bigg/\binom{N}{n}$$

$$\alpha_{12} = 1-\left\{\binom{100-3}{3}+\binom{100-1}{3}-\binom{100-3-1}{3}\right\}\bigg/\binom{100}{3} = 0.0018$$

$$\alpha_{13} = 1 - \left\{ \binom{100-3}{3} + \binom{100-4}{3} - \binom{100-3-4}{3} \right\} \Big/ \binom{100}{3} = 0.0071$$

$$\alpha_{23} = 1 - \left\{ \binom{100-1}{3} + \binom{100-4}{3} - \binom{100-1-4}{3} \right\} \Big/ \binom{100}{3} = 0.0024$$

ดังนั้น

$$\overline{y}^{*}_{HT} = \frac{1}{N}\sum_{k=1}^{\nu}\frac{u_{y_k}}{\alpha_k} = \frac{1}{100}\sum_{k=1}^{3}\frac{u_{y_k}}{\alpha_k} = \frac{1}{100}\left[\left(\frac{21}{0.0882}\right)+\left(\frac{0}{0.0300}\right)+\left(\frac{26}{0.1164}\right)\right] = 4.6277$$

และค่าประมาณความแปรปรวนของการประมาณจำนวนประชากรกวางผาโดยเฉลี่ยมีค่าเท่ากับ

$$v(\overline{y}^{*}_{HT}) = \frac{1}{N^2}\left[\sum_{k=1}^{\nu}\frac{(u_{y_k})^2}{\alpha_k}\left(\frac{1}{\alpha_k}-1\right) + 2\sum_{k=1}^{\nu-1}\sum_{l=k+1}^{\nu}\frac{u_{y_k}u_{y_h}}{\alpha_{kh}}\left(\frac{\alpha_{kh}}{\alpha_k\alpha_h}-1\right)\right]$$

$$= \frac{1}{100^2}\left[\begin{array}{l} \left(\frac{(21)^2}{0.0882}\left(\frac{1}{0.0882}-1\right)+\frac{(0)^2}{0.0300}\left(\frac{1}{0.0300}-1\right)+\frac{(26)^2}{0.1164}\left(\frac{1}{0.1164}-1\right)\right) \\ +2\left(\frac{(21)(0)}{0.0018}\right)\left(\frac{0.0018}{(0.0882)(0.0300)}-1\right)+2\left(\frac{(21)(26)}{0.0071}\right)\left(\frac{0.0071}{(0.0882)(0.1164)}-1\right) \\ +2\left(\frac{(0)(26)}{0.0024}\right)\left(\frac{0.0024}{(0.0300)(0.1164)}-1\right) \end{array}\right]$$

$$= 4.8338$$

วิธีการประมาณค่าแบบอัตราส่วน

การประมาณจำนวนประชากรกวางผาโดยเฉลี่ยด้วยวิธีการประมาณค่าแบบอัตราส่วนมีค่าเท่ากับ

$$\overline{y}_{R.HT} = \frac{\overline{y}^{*}_{HT}}{\overline{x}^{*}_{HT}}\mu_x \text{ โดยที่ } \overline{y}^{*}_{HT} = \frac{1}{N}\sum_{k=1}^{\upsilon}\frac{u_{y_k}}{\alpha_k} \text{ , } \overline{x}^{*}_{HT} = \frac{1}{N}\sum_{k=1}^{\upsilon}\frac{u_{x_k}}{\alpha_k} \text{ , } \mu_x = \frac{1}{N}\sum_{i=1}^{N}x_i$$

และ

$$\overline{y}^{*}_{HT} = \frac{1}{N}\sum_{k=1}^{\nu}\frac{u_{y_k}}{\alpha_k} = \frac{1}{100}\sum_{k=1}^{3}\frac{u_{y_k}}{\alpha_k} = \frac{1}{100}\left[\left(\frac{21}{0.0882}\right)+\left(\frac{0}{0.0300}\right)+\left(\frac{26}{0.1164}\right)\right] = 4.6277$$

$$\overline{x}^{*}_{HT} = \frac{1}{N}\sum_{k=1}^{\nu}\frac{u_{x_k}}{\alpha_k} = \frac{1}{100}\sum_{k=1}^{3}\frac{u_{x_k}}{\alpha_k} = \frac{1}{100}\left[\left(\frac{311}{0.0882}\right)+\left(\frac{0}{0.0300}\right)+\left(\frac{396}{0.1164}\right)\right] = 69.2814$$

$$\mu_x = \frac{1}{100}\sum_{i=1}^{100}(0+0+....+0) = 18.1500$$

ดังนั้น

$$\overline{y}_{R.HT} = \frac{\overline{y}^{*}_{HT}}{\overline{x}^{*}_{HT}}\mu_x = \frac{4.6277}{69.2814}18.15 = 1.2123$$

และค่าประมาณความแปรปรวนของการประมาณจำนวนประชากรกวางผาโดยเฉลี่ยมีค่าเท่ากับ

$$v(\overline{y}_{R.HT}) = \frac{1}{N^2}\left[\sum_{k=1}^{\nu}\frac{\left(u'_{y_k}\right)^2}{\alpha_k}\left(\frac{1}{\alpha_k}-1\right)+2\sum_{k=1}^{\nu-1}\sum_{l=k+1}^{\nu}\frac{u'_{y_k}u'_{y_h}}{\alpha_{kh}}\left(\frac{\alpha_{kh}}{\alpha_k\alpha_h}-1\right)\right]$$

เมื่อ

$u'_{y_k} = u_{y_k} - \hat{R}_{ACS}u_{x_k}$ โดยที่ $\hat{R}_{ACS} = \dfrac{\overline{y}^{*}_{HT}}{\overline{x}^{*}_{HT}}$

และ

$$\hat{R}_{ACS} = \frac{4.6277}{69.2814} = 0.0668$$

$$u'_{y_1} = 21-(0.0668\times 311) = 0.2252$$

$$u'_{y_2} = 0-(0.0668\times 0) = 0$$

$$u'_{y_3} = 26-(0.0668\times 396) = -0.4528$$

ดังนั้น

$$v(\bar{y}_{R.HT})=\frac{1}{100^2}\left[\begin{array}{l}\left(\frac{(0.2252)^2}{0.0882}\left(\frac{1}{0.0882}-1\right)+\frac{(0)^2}{0.0300}\left(\frac{1}{0.0300}-1\right)+\frac{(-0.4528)^2}{0.1164}\left(\frac{1}{0.1164}-1\right)\right)\\ +2\left(\frac{(0.2252)(0)}{0.0018}\right)\left(\frac{0.0018}{(0.0882)(0.0300)}-1\right)+2\left(\frac{(0.2252)(-0.4528)}{0.0071}\right)\left(\frac{0.0071}{(0.0882)(0.1164)}-1\right)\\ +2\left(\frac{(0)(-0.4528)}{0.0024}\right)\left(\frac{0.0024}{(0.0300)(0.1164)}-1\right)\end{array}\right]$$

$= 0.0028$

วิธีการประมาณค่าแบบอัตราส่วนที่ปรับปรุงโดยวิธีการ Rao-Blackwell

การประมาณจำนวนประชากรกวางผาโดยเฉลี่ยด้วยวิธีการประมาณค่าแบบอัตราส่วนมีค่าเท่ากับ

$$\bar{y}^{+}_{R.HT}=\frac{\bar{y}^{+}_{HT}}{\bar{x}^{+}_{HT}}\mu_x$$

โดยที่

$$\bar{y}^{+}_{HT}=\frac{1}{N}\sum_{k=1}^{\upsilon}\frac{m_k.w^{+}_{y_k}}{\alpha_k}\ ,\ \bar{x}^{+}_{HT}=\frac{1}{N}\sum_{k=1}^{\upsilon}\frac{m_k.w^{+}_{x_k}}{\alpha_k}\ ,\ \mu_x=\frac{1}{N}\sum_{i=1}^{N}x_i$$

เมื่อ

$$w^{+}_{y_k}=\begin{cases} w_{y_k} \\ \bar{y}_e=\frac{\sum_{i\in s} y_i}{e_s}\end{cases}\ ,\ w_{y_k}=\frac{1}{m_k}\sum y_i$$

และ

$$\bar{y}^{+}_{HT}=\frac{1}{N}\sum_{k=1}^{\nu}\frac{m_k.w^{+}_{y_k}}{\alpha_k}=\frac{1}{100}\sum_{k=1}^{3}\frac{m_k.w^{+}_{y_k}}{\alpha_k}=\frac{1}{100}\left[\left(\frac{3\times\frac{21}{3}}{0.0882}\right)+\left(\frac{1\times\frac{0}{1}}{0.0300}\right)+\left(\frac{4\times\frac{26}{4}}{0.1164}\right)\right]=4.6146$$

$$\bar{x}_{HT}^{+} = \frac{1}{N}\sum_{k=1}^{v}\frac{m_k.w_{x_k}^{+}}{\alpha_k} = \frac{1}{100}\sum_{k=1}^{3}\frac{m_k.w_{x_k}^{+}}{\alpha_k} = \frac{1}{100}\left[\left(\frac{3\times\frac{311}{3}}{0.0882}\right)+\left(\frac{1\times\frac{0}{1}}{0.0300}\right)+\left(\frac{4\times\frac{396}{4}}{0.1164}\right)\right] = 69.2813$$

$$\mu_x = \frac{1}{100}\sum_{i=1}^{100}(0+0+....+0) = 18.1500$$

ดังนั้น

$$\bar{y}_{R.HT}^{+} = \frac{4.6146}{69.2813}18.15 = 1.2089$$

และค่าประมาณความแปรปรวนของการประมาณจำนวนประชากรกวางผาโดยเฉลี่ยมีค่าเท่ากับ

$$v(\bar{y}_{R.HT}^{+}) \quad = v(\bar{y}_{R.HT}) - (\bar{y}_{R.HT}^{+} - \bar{y}_{R.HT})^2$$

$$= 0.0028 - (1.2089 - 1.2123)^2$$
$$= 0.0027$$

งานวิจัยที่เกี่ยวข้อง

ศิริประภา มโนมัธย์ (2539) ทำการศึกษาเรื่อง ประสิทธิภาพของแผนการสุ่มตัวอย่างแบบกลุ่มปรับเมื่อตัวอย่างขั้นต้นใช้วิธีการสุ่มแบบง่าย แบบมีชั้นภูมิ และแบบมีระบบ ภายใต้แบบจำลองประชากรที่กำหนด โดยใช้ร้อยละของอัตราส่วนความแปรปรวนของค่าเฉลี่ย (%eff) ระหว่างการสุ่มตัวอย่างขั้นต้นแบบมีชั้นภูมิ และแบบมีระบบเทียบกับการสุ่มตัวอย่างขั้นต้นแบบง่ายของ S.K. Thompson (1990, 1991a, 1991b) สำหรับตัวประมาณค่าเฉลี่ยที่ดัดแปลงมาจากตัวประมาณ Hansen-Hurwitz และสำหรับตัวประมาณค่าเฉลี่ยที่ดัดแปลงมาจากตัวประมาณ Horvitz-Thompson การดำเนินงานวิจัยประกอบด้วยการจำลองแบบประชากรที่มีลักษณะหายากสร้างขึ้นทั้งหมด 5 ประชากร แต่ละกรณีเป็นประชากรที่มีลักษณะต่างกันโดยใช้กระบวนการปัวซงคลัสเตอร์ ซึ่งตำแหน่ง และจำนวนจุดหลักสร้างด้วยกระบวนการปัวซงที่มีพารามิเตอร์ λ เป็น 30 ส่วนตำแหน่งของบริวารสร้างจากตัวแปรสุ่มที่มีการแจกแจงแบบปกติสองตัวแปร (Bivariate Normal) ขนาดตัวอย่างที่ใช้ในแต่ละวิธีคือ 4, 8, 16, 32 และ 64 หน่วย กระบวนการสุ่มตัวอย่างขั้นต้นใช้แบบง่าย แบบมีชั้นภูมิ และแบบมีระบบ การแบ่งชั้นภูมิแบ่งเป็น 4 ชั้นภูมิที่มีขนาดเท่า ๆ กัน ส่วนการสุ่มขั้นต้นแบบมีระบบ ใช้แผนการเลือกตัวอย่างที่มีหน่วยปฐมภูมิเป็นพื้นที่ 5×5

ตารางหน่วย และมีหน่วยทุติยภูมิเป็น 4 หน่วย ผลการวิจัยพบว่า ในการเปรียบเทียบ %eff สำหรับตัวประมาณค่าเฉลี่ยที่ดัดแปลงมาจากตัวประมาณ Hansen-Hurwitz วิธีการสุ่มขั้นต้นแบบมีระบบจะมีประสิทธิภาพในแง่ของความแม่นยำดีที่สุด ในทุกขนาดตัวอย่าง ตามลำดับรองลงมาคือการสุ่มขั้นต้นแบบมีชั้นภูมิ ยกเว้นกรณีที่ประชากร เป็นกลุ่มเล็กมีกลุ่มเดียว และเมื่อแบ่งเป็นชั้นภูมิแล้วจะมีสิ่งที่สนใจปรากฏอยู่เพียงชั้นภูมิเดียว ที่ทำให้ประสิทธิภาพของการสุ่มขั้นต้นแบบมีระบบเท่ากับการสุ่มขั้นต้นแบบมีชั้นภูมิสำหรับตัวประมาณค่าเฉลี่ยที่ดัดแปลงมาจากตัวประมาณ Horvitz-Thompson และในทุกกรณีพบว่า แผนการสุ่มตัวอย่างแบบกลุ่มปรับ สามารถใช้ได้กับประชากรที่มีลักษณะที่หายาก และอยู่รวมกันเป็นกลุ่ม ประสิทธิภาพของแผนแบบแตกต่างกันขึ้นอยู่กับลักษณะการกระจายของประชากรที่ทำการศึกษา วิธีการสุ่มตัวอย่างขั้นต้น ตัวประมาณที่ใช้ และขนาดตัวอย่างที่ทำการสุ่มในขั้นต้น และยังพบว่าประสิทธิภาพของแผนการสุ่มตัวอย่างแบบกลุ่มปรับ ที่มีการสุ่มตัวอย่างขั้นต้นแบบมีระบบใช้ได้ดีกว่าแบบอื่น ๆ รองลงมาคือแบบมีชั้นภูมิและแบบง่ายตามลำดับ

วิชาญ โชควิวัฒน์ (2546) ทำการศึกษาเรื่อง กรอบแนวคิดของแผนการสุ่มตัวอย่างแบบกลุ่มปรับ การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อศึกษากรอบแนวคิดของแผนการสุ่มตัวอย่างแบบกลุ่มปรับ และศึกษาจากแนวคิดการสุ่มตัวอย่างแบบง่ายโดยไม่ใส่คืน ภายใต้แผนแบบการสุ่มตัวอย่างแบบกลุ่มปรับของ Steven K. Thompson (1990) สำหรับตัวประมาณค่าเฉลี่ยที่ดัดแปลงมาจากตัวประมาณ Horvitz-Thompson รวมทั้งเปรียบเทียบประสิทธิภาพของตัวประมาณค่าเฉลี่ยที่ทำการสุ่มตัวอย่างขั้นต้นอย่างง่ายโดยไม่ใส่คืนภายใต้แผนการสุ่มตัวอย่างแบบกลุ่มปรับ และแผนการสุ่มตัวอย่างที่ยังไม่ได้ปรับ ด้วยค่าประสิทธิภาพสัมพันธ์ โดยแบ่งพื้นที่ศึกษาเป็น 100 หน่วย ประชากรที่ใช้สร้างมาจากกระบวนการปัวซงคลัสเตอร์ ซึ่งตำแหน่ง และจำนวนของจุดสร้างจากกระบวนการปัวซงที่มีพารามิเตอร์ λ เป็น 20 โดยจำลองประชากรเป็น 3 กรณี และตำแหน่งของบริวารต่าง ๆ สร้างจากตัวแปรสุ่มที่มีการแจกแจงแบบปกติ 2 ตัวแปร (Bivariate Normal) ขนาดตัวอย่างคือ 4, 8, 16 และ 32 หน่วย ซึ่งแต่ละหน่วยตัวอย่างแบ่งเป็น 100 รูปแบบ ผลการวิจัยพบว่าเมื่อเปรียบเทียบประสิทธิภาพในแง่ของความแม่นยำของตัวประมาณค่าเฉลี่ยที่ทำการสุ่มตัวอย่างขั้นต้นอย่างง่ายโดยไม่ใส่คืนภายใต้แผนการสุ่มตัวอย่างแบบกลุ่มปรับกับตัวประมาณค่าเฉลี่ยจากแผนการสุ่มตัวอย่างแบบที่ไม่ปรับ พบว่าตัวประมาณค่าเฉลี่ยแบบกลุ่มปรับมีประสิทธิภาพในแง่ความแม่นยำมากกว่าตัวประมาณค่าเฉลี่ยจากแผนการสุ่มตัวอย่างแบบที่ไม่ปรับทั้ง 3 กรณี และทุก ๆ ขนาดตัวอย่างที่เพิ่มขึ้นจะทำให้ตัวประมาณค่าเฉลี่ยของทั้ง 2 แบบมีความแม่นยำมากขึ้นและการสุ่มตัวอย่างแบบกลุ่มปรับเหมาะสมกับการสุ่มตัวอย่างของสิ่งตัวอย่างที่เราสนใจที่อยู่เกาะ

กลุ่มกัน ซึ่งเป็นลักษณะของสิ่งที่หายาก เช่น ฝูงสัตว์ พืชต่าง ๆ ฟอสซิล รวมทั้งแร่ธาตุ เป็นต้น ส่วนการสุ่มตัวอย่างแบบที่ไม่ปรับเหมาะสมกับการสุ่มตัวอย่างของสิ่งที่เราสนใจ ซึ่งมีลักษณะที่อยู่กระจายตัวกัน

Thompson (1990) ทำการศึกษา และเปรียบเทียบประสิทธิภาพของแผนการสุ่มตัวอย่างแบบกลุ่มปรับกับแผนการสุ่มตัวอย่างแบบกลุ่มธรรมดาที่ยังไม่มีการปรับกลุ่ม และเปรียบเทียบประสิทธิภาพตัวประมาณค่าเฉลี่ยที่ดัดแปลงมาจากตัวประมาณ Hansen-Hurwitz และ Horvitz-Thompson ที่หน่วยตัวอย่างใช้วิธีการสุ่มอย่างง่าย เกณฑ์ที่ใช้ในการเปรียบเทียบคือค่าความคลาดเคลื่อนกำลังสองเฉลี่ย จากการวิจัยพบว่า ถ้าใช้ตัวประมาณของแผนการสุ่มตัวอย่างแบบกลุ่มธรรมดาที่ยังไม่มีการปรับกลุ่มนั้นมาประมาณค่าเฉลี่ยสิ่งที่สนใจที่มีลักษณะหายาก และอยู่รวมกันเป็นแล้วนั้น จะทำให้เกิดความเอนเอียง และพบว่าตัวประมาณค่าเฉลี่ยที่ดัดแปลงมาจากตัวประมาณ Horvitz-Thompson มีค่าความแปรปรวนต่ำกว่าตัวประมาณค่าเฉลี่ยที่ดัดแปลงมาจากตัวประมาณ Hansen-Hurwitz

Dryver and Thompson (2005) ทำการศึกษา และเปรียบเทียบประสิทธิภาพตัวประมาณค่าเฉลี่ยที่ปรับปรุงโดยวิธีการ Rao-Blackwell กับตัวประมาณค่าเฉลี่ยแบบพื้นฐานที่ตัวประมาณค่าเฉลี่ยดัดแปลงมาจากตัวประมาณ Hansen-Hurwitz และ Horvitz-Thompson ผลการวิจัยพบว่า ตัวประมาณค่าเฉลี่ยดัดแปลงมาจากตัวประมาณ Hansen-Hurwitz และ Horvitz-Thompson ที่ปรับปรุงโดยวิธีการ Rao-Blackwell จะมีประสิทธิภาพที่ดีกว่าตัวประมาณค่าเฉลี่ยแบบพื้นฐาน และตัวประมาณค่าเฉลี่ยที่ดัดแปลงมาจากตัวประมาณ Horvitz-Thompson ที่ปรับปรุงโดยวิธีการ Rao-Blackwell และตัวประมาณค่าเฉลี่ยแบบพื้นฐานจะมีประสิทธิภาพดีกว่าตัวประมาณค่าเฉลี่ยที่ดัดแปลงมาจากตัวประมาณ Hansen-Hurwitz

Chao, Lin and Chiang (2008) ทำการศึกษา และเปรียบเทียบประสิทธิภาพตัวประมาณค่าเฉลี่ยแบบอัตราส่วนที่ปรับปรุงโดยวิธีการ Rao-Blackwell ที่ตัวประมาณค่าเฉลี่ยดัดแปลงมาจากตัวประมาณ Hansen-Hurwitz และ Horvitz-Thompson กับตัวประมาณอัตราส่วนแบบพื้นฐาน ที่ Dryver and Chao (2007) เป็นผู้ศึกษาไว้ ผลการวิจัยพบว่า ตัวประมาณค่าเฉลี่ยแบบอัตราส่วนที่ปรับปรุงโดยวิธีการ Rao-Blackwell จะมีฟังก์ชันของสถิติพอเพียงต่ำที่สุด (Minimal Sufficient Statistic) และมีประสิทธิภาพมากกว่าตัวประมาณค่าเฉลี่ยอัตราส่วนแบบพื้นฐาน