รายงานผลการวิจัยฉบับสมบูรณ์
การเปรียบเทียบประสิทธิภาพวิธีประมาณค่าเฉลี่ยภายใต้แผนการสุ่มตัวอย่างแบบกลุ่มปรับด้วยการสุ่มหน่วยตัวอย่างแบบแบ่งชั้นภูมิ

วิไลวรรณ รัตนกูล

งานวิจัยนี้ได้รับทุนอุดหนุนการวิจัยจากงบประมาณแผ่นดิน

มหาวิทยาลัยราชภัฏอุตรดิตถ์

ประจำปีงบประมาณ 2561

พ.ศ.2561

กิตติกรรมประกาศ

ขอขอบพระคุณทุนอุดหนุนการวิจัยจากสำนักงานคณะกรรมการวิจัยแห่งชาติ (วช.) ปีงบประมาณ 2561 มหาวิทยาลัยราชภัฏอุตรดิตถ์ ที่ให้การสนับสนุนทุนการทำวิจัย ทำให้งานวิจัยชิ้นนี้สำเร็จลุล่วง

ขอขอบพระคุณ ผู้ช่วยศาสตราจารย์ ดร.ทวีศักดิ์ ศิริพรไพบูลย์ ที่ปรึกษางานวิจัย ที่กรุณาให้คำปรึกษา พร้อมทั้งให้คำแนะนำ ตลอดจนช่วยเหลือแก้ไขข้อบกพร่องต่าง ๆ ด้วยดีตลอดมา จนกระทั่งวิจัยฉบับนี้เสร็จสมบูรณ์ ผู้วิจัยรู้สึกซาบซึ้ง และขอกราบขอบพระคุณเป็นอย่างสูงมา ณ โอกาสนี้

ขอกราบขอบพระคุณ ผู้ช่วยศาสตราจารย์ ดร.เกตุจันทร์ จำปาไชยศรี ที่กรุณาให้คำแนะนำตรวจแก้ไขข้อบกพร่องวิจัยฉบับนี้ให้สมบูรณ์มากยิ่งขึ้น จึงขอกราบขอบพระคุณเป็นอย่างสูง

ขอขอบพระคุณหลักสูตรคณิตศาสตร์ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏอุตรดิตถ์ที่ให้ความอนุเคราะห์ในด้านต่างๆ ตลอดระยะเวลาในการทำวิจัยนี้

เหนือสิ่งอื่นใดขอกราบขอบพระคุณ บิดา มารดา และสมาชิกทุกคนในครอบครัวของผู้วิจัยที่ให้กำลังใจและให้การสนับสนุนในทุก ๆ ด้านอย่างดีที่สุดเสมอมา

วิไลวรรณ รัตนกูล
สิงหาคม 2561

ชื่อโครงการวิจัย การเปรียบเทียบประสิทธิภาพวิธีประมาณค่าเฉลี่ยภายใต้แผนการสุ่มตัวอย่างแบบกลุ่มปรับด้วยการสุ่มหน่วยตัวอย่างแบบแบ่งชั้นภูมิ The Efficiency Comparison of Mean Estimator Methods under Adaptive Cluster Sampling With Stratified Random Unit Selection)

ชื่อผู้วิจัย วิไลวรรณ รัตนกูล หัวหน้าโครงการ

บทคัดย่อ

การวิจัยนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพวิธีประมาณค่าเฉลี่ยภายใต้แผนการสุ่มตัวอย่างแบบกลุ่มปรับ 3 วิธีคือ วิธีประมาณค่าอย่างง่าย วิธีประมาณค่าแบบอัตราส่วน และวิธีประมาณค่าแบบอัตราส่วนที่ปรับปรุงโดยวิธีการ Rao-Blackwell เมื่อสุ่มหน่วยตัวอย่างแบบแบ่งชั้นภูมิ โดยกำหนดสถานการณ์ 2 กรณี คือ กรณีหน่วยขอบแยกกัน และกรณีหน่วยขอบร่วมกัน กำหนดขนาดพื้นที่ย่อยเท่ากับ 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 16, 20 และ 24 หน่วย แบ่งความสัมพันธ์ 3 ระดับ คือ ระดับต่ำ (0, 0.4) ปานกลาง [0.4, 0.7) และสูง [0.7, 1.00) ทำการจำลองข้อมูล ในแต่ละสถานการณ์ทำซ้ำ 1,000 ครั้ง โดยใช้ความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Square Error: MSE) เป็นเกณฑ์ในการเปรียบเทียบ โดยวิธีประมาณค่าที่มีค่าความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุดจะเป็นวิธีประมาณค่าที่มีประสิทธิภาพมากที่สุด

ผลการวิจัย กรณีประมาณค่าเฉลี่ยสำหรับประชากรกรณีหน่วยขอบแยกกันและหน่วยขอบร่วมกัน พบว่าเมื่อพิจารณาทุกขนาดพื้นที่ย่อย วิธีการประมาณค่าอย่างง่ายมีค่าความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุดเป็นส่วนใหญ่เมื่อความสัมพันธ์อยู่ในระดับต่ำ และวิธีการประมาณค่าแบบอัตราส่วนที่ปรับปรุงโดยวิธีการ Rao-Blackwell มีค่าความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุดเป็นส่วนใหญ่ เมื่อความสัมพันธ์อยู่ในระดับปานกลางและสูง ยกเว้นกรณีหน่วยขอบแยกกัน 3 กลุ่มที่ระดับความสัมพันธ์ปานกลางพบว่า วิธีการประมาณค่าแบบอัตราส่วนมีค่าความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุดเป็นส่วนใหญ่

Title The Efficiency Comparison of Mean Estimator Methods under Adaptive Cluster Sampling With Stratified Random Unit Selection

Author Wilaiwan Rattanakul

ABSTRACT

The objective of this research was to compare the efficiency among three mean estimation methods: Simple Estimation Methods (SE), Ratio Estimation Methods (RA) and Improved Ratio Estimation Methods by Rao-Blackwell (RB) under adaptive cluster sampling, when using Stratified Random Unit Selection. Simulations were used to study into two cases: separate edge unit case and combined edge unit case. There were 13 sample sizes: 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 16, 20 and 24. Three levels of correlation coefficients were used: low (0, 0.4), medium [0.4, 0.7) and high [0.7, 1.00). In each situation, data was simulated and repeated 1,000 times. The mean square error was used as a criterion for comparison.

The results indicated that both of two cases give the same results. When correlation coefficient between separate edge unit case and combined edge unit case is low, simple estimation methods gives minimum mean square error. Despite the correlation coefficient is medium on high, improved ratio estimation methods by Rao-Blackwell gives minimum mean square error except when correlation coefficient for combined edge unit case with 3 clusters is medium, ratio estimation methods gives minimum mean square error.

สารบัญ

สารบัญ (ต่อ)

บทที่ **หน้า**

สารบัญภาพ

สารบัญตาราง

ตาราง **หน้า**

บทที่ 1
บทนำ

ความเป็นมาของปัญหา

เทคนิคการสุ่มตัวอย่างเป็นระเบียบวิธีการทางสถิติอย่างหนึ่งที่นับว่ามีความสำคัญอย่างยิ่ง สามารถนำไปใช้ประโยชน์ได้ทุกหน่วยงานไม่ว่าจะเป็นหน่วยงานในภาคราชการ ภาครัฐวิสาหกิจ ภาคเอกชน องค์กรระหว่างประเทศหรือในแวดวงการวิชาการ และสื่อมวลชน เป็นต้น ดังเช่นในสำนักงานสถิติแห่งชาติได้ดำเนินการสำรวจการทำงานของประชาชนคนไทย เพื่อเป็นข้อมูลพื้นฐานในการแสดงภาวะการมีงานทำ อัตราการว่างงาน หรือโครงการสำรวจความคิดเห็นของประชาชนเกี่ยวกับสถานการณ์การแพร่ระบาดยาเสพติด (สำนักงานสถิติแห่งชาติ, 2554) และโครงการสำรวจทรัพยากรธรณีในทะเล (กรมทรัพยากรธรณี, 2550) เป็นต้น ทั้งนี้โครงการสำรวจด้วยตัวอย่างจะมีจุดเด่นที่ดีกว่าการสำรวจข้อมูลที่ใช้ทั้งประชากรหรือที่เรียกว่า การทำสำมะโน เนื่องจาก 1) เสียงบประมาณน้อยกว่า 2) ใช้เวลาน้อยกว่า 3) ใช้อัตรากำลังคนน้อยกว่า และ 4) การศึกษาจะมีคุณภาพที่ดีกว่า (สุชาดา กีระนันทน์, 2542, หน้า 5)

การสำรวจด้วยตัวอย่างเป็นวิธีการคัดเลือกหน่วยตัวอย่างเพียงบางส่วนจากประชากร และการที่จะได้หน่วยตัวอย่างที่เป็นตัวแทนที่ดีของประชากรจะต้องพิจารณาความเหมาะสมของคุณลักษณะประชากร โดยการเลือกใช้แผนการสุ่มตัวอย่างที่ดี ควรรู้จักกับโครงสร้างของประชากร ซึ่งจะทำให้เลือกแผนการสุ่มตัวอย่างได้เหมาะสม นอกจากนี้การใช้วิธีการประมาณค่าที่เหมาะสมกับสถานการณ์ที่สนใจศึกษาจะทำให้สามารถผลิตสถิติที่มีคุณภาพมีวิธีการสุ่มตัวอย่างหลายวิธีที่สามารถประยุกต์ใช้กับประชากรที่หายาก เช่น การสุ่มตัวอย่างแบบชั้นภูมิ (Stratified Sampling) การสุ่มตัวอย่างแบบสองขั้น (Two Phase Sampling) การสุ่มตัวอย่างแบบลูกหิมะ (Snowball Sampling) การสุ่มจัวอย่างแบบข่ายงาน () โครงการสำรวจภาวะการมีงานทำของประเทศไทย ได้ใช้แผนการสุ่มตัวอย่างแบบมีชั้นภูมิ 3 ขั้นตอน (Stratified Three-stage Cluster Sampling) และใช้ตัวประมาณแบบอัตราส่วนในการประมาณค่า (สำนักงานสถิติแห่งชาติ, 2542) ซึ่งวิธีดังกล่าวเป็นการวางแผนที่ประสบความสำเร็จ และสถิติที่ได้จะมีความน่าเชื่อถือ ซึ่งแบบแผนนี้ใช้กันอย่างกว้างขวางทั่วโลก แต่แบบแผนนี้อาจไม่ประสบความสำเร็จเมื่อนำไปใช้กับสิ่งที่สนใจศึกษาที่มีลักษณะหายาก และอยู่รวมกันเป็นกลุ่ม ดังนั้นจึงมีการปรับแผนการสุ่มตัวอย่างให้เหมาะสมกับสถานการณ์มากยิ่งขึ้น (ประชุม สุวัตถี, 2552, หน้า 445)

แผนการสุ่มตัวอย่างแบบกลุ่มปรับ (Adaptive Cluster Sampling) เป็นวิธีการที่ดัดแปลงมาจากแผนการสุ่มตัวอย่างแบบกลุ่ม (Cluster Sampling) เพื่อใช้สำรวจสิ่งที่หายากที่อยู่รวมกันเป็นกลุ่ม เช่น ต้องการประมาณจำนวนประชากรสัตว์ป่าที่หายากในพื้นที่แห่งหนึ่ง โดยที่หน่วยตัวอย่างถูกสุ่มด้วยแผนการสุ่มตัวอย่างความน่าจะเป็นขั้นพื้นฐาน (Conventional Probability Sampling) เช่น การสุ่มตัวอย่างแบบง่าย การสุ่มตัวอย่างแบบมีชั้นภูมิ และการสุ่มตัวอย่างแบบมีระบบ เป็นต้น แล้วอาจจะไม่พบสัตว์ป่า หรืออาจจะพบสัตว์ป่าที่สนใจเพียงหน่วยเดียวหรือว่าน้อยมาก โดยข้อมูลที่ได้อาจจะไม่เพียงพอต่อการคำนวณ และอาจส่งผลให้ค่าประมาณที่ได้ไม่เสถียรภาพพอ ซึ่งอาจก่อให้เกิด

ความเอนเอียงในการประมาณค่าได้ ดังนั้นการใช้แผนการสุ่มตัวอย่างแบบกลุ่มปรับเป็นแนวทางหนึ่งที่นิยมใช้ในการแก้ไขปัญหาดังกล่าว (Thompson, 1990, pp.1050-1059) โดยที่เมื่อใดก็ตามที่เราทำการสุ่มตัวอย่างแล้วพบสัตว์ป่าก็จะทำการพิจารณาหน่วยบริเวณใกล้เคียงถ้าเป็นไปตามเงื่อนไขก็จะทำการขยายหน่วยถัดไปทางซ้าย ขวา บน และล่าง ทำเช่นนี้ไปเรื่อย ๆ จนกระทั่งไม่พบสัตว์ป่าที่สนใจ ข้อมูลทั้งหมดที่พบก็จะเพียงพอในการนำไปคำนวณ การพบสิ่งที่สนใจศึกษาโดยบังเอิญ และการเพิ่มตัวอย่างในหน่วยที่ใกล้เคียง จะทำให้มีอัตราการพบสิ่งที่ศึกษาในอัตราที่ค่อนข้างสูง (สุชาดา กีระนันทน์, 2542, หน้า 352) ดังนั้นจะเห็นได้ว่าแผนการสุ่มตัวอย่างแบบกลุ่มปรับเป็นแผนแบบการสุ่มตัวอย่างที่น่าสนใจ และเป็นประโยชน์อย่างยิ่งถ้านำไปใช้ให้เหมาะสมกับลักษณะประชากรที่มีลักษณะอยู่รวมกันเป็นกลุ่ม ๆ โดยผู้วิจัยได้ทำการรวบรวม และสรุปผลงานวิจัยต่าง ๆ ที่เกี่ยวข้องกับแผนการสุ่มตัวอย่างแบบกลุ่มปรับ และสามารถสรุปแนวคิดได้ดังนี้

ศิรประภา มโนมัธย์ (2539) ได้ทำการศึกษาและเปรียบเทียบประสิทธิภาพของแผนการสุ่มตัวอย่างแบบกลุ่มปรับเมื่อตัวอย่างขั้นต้นใช้วิธีการสุ่มแบบง่าย แบบมีชั้นภูมิ และแบบมีระบบ โดยใช้ร้อยละของอัตราส่วนความแปรปรวนของค่าเฉลี่ย เป็นเกณฑ์ในการเปรียบเทียบ จากการวิจัยพบว่าประสิทธิภาพของแผนการสุ่มตัวอย่างแบบมีระบบมีประสิทธิภาพดีที่สุด รองลงมาคือ แผนการสุ่มตัวอย่างแบบมีชั้นภูมิและแผนการสุ่มตัวอย่างแบบง่ายตามลำดับ วิชาญ โชควิวัฒน์ (2546) และ Thompson (1990) ได้ทำการศึกษาและเปรียบเทียบประสิทธิภาพตัวประมาณค่าเฉลี่ยภายใต้แผนการสุ่มตัวอย่างแบบกลุ่มปรับและแผนการสุ่มตัวอย่างแบบกลุ่มที่ไม่ปรับ (Nonadaptive) โดยใช้ค่าประสิทธิภาพสัมพันธ์ และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยเป็นเกณฑ์ในการเปรียบเทียบตามลำดับ ซึ่งผลการวิจัยสอดคล้องไปในทิศทางเดียวกัน โดยพบว่าตัวประมาณค่าเฉลี่ยแบบกลุ่มปรับมีประสิทธิภาพที่ดีกว่าตัวประมาณค่าเฉลี่ยจากแผนการสุ่มตัวอย่างแบบกลุ่มที่ไม่ปรับ นอกจากนี้ Thompson ยังพบว่าตัวประมาณค่าเฉลี่ยที่ดัดแปลงมาจากตัวประมาณ Horvitz-Thompson มีค่าความแปรปรวนต่ำกว่าตัวประมาณค่าเฉลี่ยที่ดัดแปลงมาจากตัวประมาณ Hansen-Hurwitz นอกจากการประมาณค่าดังกล่าวแล้วได้มีผู้ที่ศึกษาวิธีประมาณค่าโดยใช้ประโยชน์จากตัวแปรช่วยและมีการพัฒนาด้วยกระบวนการ Rao-Blackwell มาทำการศึกษา โดยในปี ค.ศ. 2005 Dryver และ Thompson ได้ทำการศึกษาและเปรียบเทียบประสิทธิภาพตัวประมาณค่าเฉลี่ยที่ปรับปรุงโดยวิธีการ Rao-Blackwell กับตัวประมาณค่าเฉลี่ยแบบพื้นฐาน โดยที่ตัวประมาณค่าเฉลี่ยดัดแปลงมาจากตัวประมาณ Hansen-Hurwitz และ Horvitz-Thompson ผลการวิจัยพบว่าตัวประมาณค่าเฉลี่ย ที่ปรับปรุงโดยวิธีการ Rao-Blackwell จะมีประสิทธิภาพที่ดีกว่าตัวประมาณค่าเฉลี่ยแบบพื้นฐาน และในปี ค.ศ.2007 Dryver และ Chao มีแนวคิดในการใช้ประโยชน์จากตัวแปรช่วยในการประมาณค่าซึ่งได้ทำการศึกษาและเปรียบเทียบประสิทธิภาพตัวประมาณค่าเฉลี่ยแบบอัตราส่วนกับตัวประมาณค่าเฉลี่ยแบบพื้นฐาน โดยที่ตัวประมาณค่าเฉลี่ยดัดแปลงมาจากตัวประมาณ Hansen-Hurwitz และ Horvitz-Thompson ผลการวิจัยพบว่าตัวประมาณค่าเฉลี่ยแบบอัตราส่วน จะมีประสิทธิภาพที่ดีกว่าตัวประมาณค่าเฉลี่ย ต่อมาในปี ค.ศ.2008 Chao, Lin และ Chiang มีแนวคิดในการปรับปรุงตัวประมาณค่าเฉลี่ยแบบอัตราส่วนให้มีประสิทธิภาพเพิ่มมากขึ้น โดยใช้ทฤษฎี Rao-Blackwell มาทำการศึกษาและเปรียบเทียบประสิทธิภาพตัวประมาณค่าเฉลี่ยแบบอัตราส่วนที่ปรับปรุงโดยวิธีการ Rao-Blackwell กับตัวประมาณค่าเฉลี่ยแบบอัตราส่วน ที่ Dryver และ Chao ศึกษาในปี ค.ศ. 2007

โดยที่ตัวประมาณค่าเฉลี่ยดัดแปลงมาจากตัวประมาณ Hansen-Hurwitz และ Horvitz-Thompson และผลการวิจัยพบว่า ตัวประมาณค่าเฉลี่ยแบบอัตราส่วนที่ปรับปรุงโดยวิธีการ Rao-Blackwell จะเป็นฟังก์ชันของสถิติพอเพียงต่ำที่สุด (Minimal Sufficient Statistic) และมีประสิทธิภาพที่ดีกว่าตัวประมาณค่าเฉลี่ยแบบอัตราส่วน (ศิรประภา มโนมัธย์, 2539; Thompson, 1990; วิชาญ โชควิวัฒน์, 2546; Dryver and Thompson, 2005; Dryver and Chao, 2007; Chao, Lin and Chiang, 2008)

จากการศึกษาแนวคิดและงานวิจัยที่เกี่ยวข้องจะเห็นได้ว่าการใช้ประโยชน์จากตัวแปรช่วยที่เหมาะสม จะทำให้การประมาณค่ามีความคลาดเคลื่อนลดลง และการใช้ทฤษฎีสถิติที่มีความพอเพียงจะทำให้ได้สถิติที่มีคุณภาพและน่าเชื่อถือ ดังนั้นผู้วิจัยจึงมีความสนใจที่จะทำการเปรียบเทียบวิธีประมาณค่าเฉลี่ยด้วยวิธีประมาณค่าอย่างง่าย วิธีประมาณค่าแบบอัตราส่วน และวิธีประมาณค่าแบบอัตราส่วนที่ปรับปรุงโดยวิธีการ Rao-Blackwell โดยใช้ ตัวประมาณค่าเฉลี่ยที่ดัดแปลงมาจากตัวประมาณ Horvitz-Thompson

จุดมุ่งหมายของการศึกษา

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพวิธีประมาณค่าเฉลี่ย ภายใต้แผนการสุ่มตัวอย่างแบบกลุ่มปรับ ด้วยวิธีประมาณค่าอย่างง่าย วิธีประมาณค่าแบบอัตราส่วน และวิธีประมาณค่าแบบอัตราส่วนที่ปรับปรุงโดยวิธีการ Rao-Blackwell เมื่อสุ่มหน่วยตัวอย่างแบบแบ่งชั้นภูมิ

ขอบเขตของงานวิจัย

การวิจัยครั้งนี้ผู้วิจัยได้กำหนดขอบเขต และขั้นตอนดำเนินการวิจัย ดังนี้

1. แบ่งพื้นที่ทำการศึกษาออกเป็น 100 พื้นที่ย่อย จำนวน 2 ชั้นภูมิ ดังภาพ 1

2. จำลองแบบประชากรที่มีลักษณะหายาก และอยู่รวมกันเป็นกลุ่ม 2 กรณีที่แตกต่างกันลงใน 100 พื้นที่ย่อยโดยอาศัยกระบวนปัวซงคลัสเตอร์ ที่มีพารามิเตอร์ λ เป็น 30 โดยใช้โปรแกรม Minitab 14 ในการจำลองแบบประชากร ซึ่งลักษณะประชากร 2 กรณีประกอบด้วย

   2.1 ประชากรกรณีที่ 1 แบ่งเป็น 3 ลักษณะ
      2.1.1 หน่วยขอบแยกกัน 2 กลุ่ม
      2.1.2 หน่วยขอบแยกกัน 3 กลุ่ม
      2.1.3 หน่วยขอบแยกกัน 4 กลุ่ม
   2.2 ประชากรกรณีที่ 2 แบ่งเป็น 3 ลักษณะ
      2.2.1 หน่วยขอบร่วมกัน 2 กลุ่ม
      2.2.2 หน่วยขอบร่วมกัน 3 กลุ่ม
      2.2.3 หน่วยขอบร่วมกัน 4 กลุ่ม

3. ตัวแปรในแต่ละสถานการณ์ประกอบด้วย

   3.1 ตัวแปร Y คือ ตัวแปรที่ต้องการศึกษา เป็นข้อมูลเชิงปริมาณ
   3.2 ตัวแปร X คือ ตัวแปรช่วยของสิ่งที่ต้องการศึกษา เป็นข้อมูลเชิงปริมาณในพื้นที่ย่อย

3.3 กำหนดความสัมพันธ์ระหว่างตัวแปรเชิงปริมาณ Y กับตัวแปรเชิงปริมาณ X

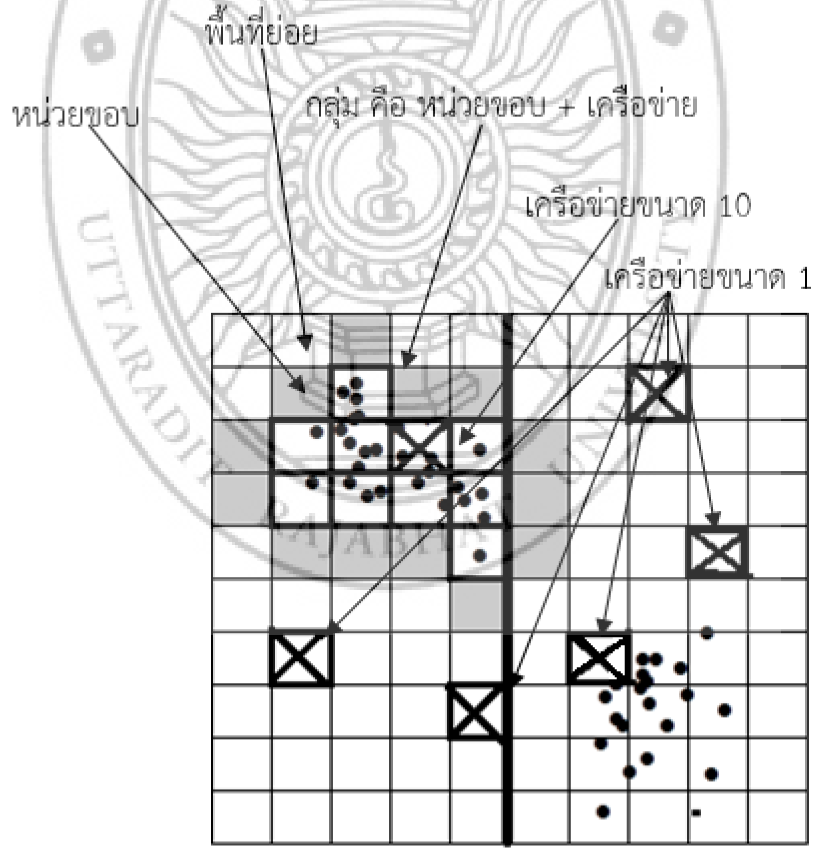
ความสัมพันธ์ระดับต่ำ	อยู่ในช่วง (0 , 0.4)
ความสัมพันธ์ระดับปานกลาง	อยู่ในช่วง [0.4, 0.7)
ความสัมพันธ์ระดับสูง	อยู่ในช่วง [0.7, 1.00)

4. กำหนดขนาดพื้นที่ย่อยที่ใช้ในการวิจัยเท่ากับ 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 16, 20 และ 24 ในแต่ละชั้นภูมิ

5. คำนวณหาค่าเฉลี่ย และความแปรปรวนของวิธีประมาณค่าเฉลี่ย ภายใต้แผนการสุ่มตัวอย่างแบบกลุ่มปรับด้วยวิธีประมาณค่าอย่างง่าย วิธีประมาณค่าแบบอัตราส่วน และวิธีประมาณค่าแบบอัตราส่วนที่ปรับปรุงโดยวิธีการ Rao-Blackwell เมื่อสุ่มหน่วยตัวอย่างแบบแบ่งชั้นภูมิ

6. เปรียบเทียบประสิทธิภาพของแต่ละวิธีโดยใช้เกณฑ์ของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย

7. ทำซ้ำด้วยโปรแกรม MATLAB 1,000 รอบ



ภาพ 1 แสดงการสุ่มหน่วยตัวอย่างด้วยแผนการสุ่มตัวอย่างแบบแบ่งชั้นภูมิที่มีขนาด 3 หน่วย

นิยามศัพท์เฉพาะ

แผนการสุ่มตัวอย่างแบบกลุ่มปรับ หมายถึง แผนการสุ่มตัวอย่างแบบที่ปรับกระบวนการสุ่มตัวอย่างให้สามารถนำลักษณะข้อมูลที่เก็บได้มากำหนดหน่วยตัวอย่างด้วยการขยายหน่วยตัวอย่างที่เป็นไปตามเงื่อนไขที่สนใจศึกษาทั้งด้าน บน ล่าง ขวา และซ้าย จนกระทั่งไม่พบหน่วยตัวอย่างที่เป็นไปตามเงื่อนไข

แผนการสุ่มตัวอย่างแบบที่ไม่ได้ปรับ หมายถึง แผนการสุ่มตัวอย่างแบบทั่วไปหรือแบบพื้นฐาน เช่น แผนการสุ่มตัวอย่างแบบง่าย แผนการสุ่มตัวอย่างแบบมีชั้นภูมิ หรือ แผนการสุ่มตัวอย่างแบบมีระบบ เป็นต้น

เงื่อนไขของสิ่งที่สนใจ หมายถึง สิ่งที่กำหนดเพื่อทำการขยายหน่วยถัดไป โดยงานวิจัยนี้ใช้เงื่อนไข $C=\{y:y\geq 3\}$ โดยที่ให้ C เป็นเซตของค่าที่มากกว่าหรือเท่ากับ 3 โดยถ้า y ซึ่งแทนสิ่งที่สนใจที่มีลักษณะหายาก และอยู่รวมกันเป็นกลุ่มที่อยู่ใน C มีค่ามากกว่าหรือเท่ากับ 3 ซึ่งเป็นไปตามเงื่อนไขที่กำหนดก็จะพิจารณาหน่วยบริเวณใกล้เคียงเข้าไปเป็นหน่วยตัวอย่าง

หน่วยขอบ (Edge Unit) หมายถึง หน่วยที่ไม่เป็นไปตามเงื่อนไขที่กำหนด แต่หน่วยที่ใกล้เคียงของหน่วยขอบนั้นสอดคล้องตามเงื่อนไข

เครือข่าย (Network) หมายถึง กลุ่มย่อยซึ่งประกอบด้วยหน่วยย่อยที่เป็นไปตามเงื่อนไขภายในกลุ่ม แต่ถ้าหน่วยใดที่ไม่เป็นไปตามเงื่อนไขแต่ถูกเลือกเป็นตัวอย่างจะให้เครือข่ายนั้นมีขนาดเป็น 1

กลุ่ม (Cluster) หมายถึง กลุ่มที่มีการเก็บหน่วยทั้งหมดที่ถูกสังเกตภายใต้แผนการสุ่มตัวอย่างแบบกลุ่มปรับของการสุ่มตัวอย่างในหน่วยที่ i ซึ่งประกอบด้วยหน่วยขอบรวมกับเครือข่าย

พื้นที่ย่อย (Plot) หมายถึง พื้นที่เล็ก ๆ ที่รวมกันเป็นประชากร

ประโยชน์ที่คาดว่าจะได้รับ

1. เพื่อทราบถึงประสิทธิภาพของ วิธีประมาณค่าเฉลี่ยทั้ง 3 วิธีการ ว่าเหมาะสมในสถานการณ์ใด

2. เพื่อเกิดความเข้าใจที่ชัดเจนในทฤษฎีการสุ่มตัวอย่างแบบกลุ่มปรับ และนำไปประยุกต์ใช้สำหรับการสำรวจตัวอย่างที่หน่วยตัวอย่างมีลักษณะที่หายาก และอยู่รวมกันเป็นกลุ่ม

บทที่ 2
เอกสาร และงานวิจัยที่เกี่ยวข้อง

ในการวิจัยครั้งนี้จะกล่าวถึงเอกสาร และงานวิจัยที่เกี่ยวข้องกับการเปรียบเทียบประสิทธิภาพวิธีประมาณค่าเฉลี่ยที่ดัดแปลงมาจากตัวประมาณ Horvitz-Thompson ภายใต้แผนการสุ่มตัวอย่างแบบกลุ่มปรับ เมื่อสุ่มหน่วยตัวอย่างแบบแบ่งชั้นภูมิ 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 เพื่อนำประชากรที่จำลองได้ไปใช้ศึกษาเกี่ยวกับความแปรปรวนของตัวประมาณค่าเฉลี่ยของแผนการสุ่มตัวอย่างแบบกลุ่มปรับที่หน่วยตัวอย่างสุ่มแบบแบ่งชั้นภูมิ โดยทฤษฎีในส่วนนี้ประกอบด้วย กระบวนการปัวซงและ 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 คือการเลือกตัวอย่างจากประชากรหนึ่งที่ตัวอย่างที่เป็นไปได้แต่ละตัวอย่างมีความน่าจะเป็นที่จะถูกเลือกเท่ากัน (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 จากนั้นสุ่มหน่วยตัวอย่างโดยการดูหมายเลขจากตารางเลขสุ่ม (ตารางเลขสุ่มเป็นตารางที่ประกอบไปด้วย ตัวเลขที่เรียงแบบไม่เจาะจงตามแถว และคอลัมน์ต่าง ๆ) โดยที่จะเริ่มที่แถวใดหรือคอลัมน์ใดก็ได้ แล้วแต่ผู้เลือกเป็นผู้กำหนด อ่านค่าไปจนครบตามจำนวนที่ต้องการ เช่น ถ้าต้องการเลือกตัวอย่างขนาด 10 จากประชากรที่มีขนาด 1,000

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

ขั้นที่ 2 จากนั้นสุ่มเลือกตัวอย่างจากตารางเลขสุ่ม โดยจะเริ่มที่แถวใดคอลัมน์ใดก็ได้ 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 นอกจากนี้เรายังสามารถใช้วิธีการเลือกหน่วยตัวอย่างในแต่ละชั้นภูมิที่แตกต่างกัน เช่นชั้นภูมิที่ 1 อาจใช้การสุ่มตัวอย่างแบบง่าย ชั้นภูมิที่ 2 อาจใช้การสุ่มตัวอย่างแบบมีระบบก็ได้ สมมติ ต้องการศึกษาการจัดการเรียนการสอนของโรงเรียนระดับประถมศึกษาในจังหวัดแห่งหนึ่งจำนวน 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

การสุ่มตัวอย่างแบบมีระบบเป็นวิธีการเลือกตัวอย่างที่มีความสะดวกในการใช้งานหลายกรณีที่การหาหรือสร้างกรอบสำหรับเลือกตัวอย่างกระทำได้ยากหรือไม่สามารถกระทำได้ด้วย

เหตุต่าง 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 และในกรณีที่จัดเก็บข้อมูลครบทุกหน่วยตัวอย่างในกลุ่มของหน่วยตัวอย่างที่สุ่มได้นั้น 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 ก็ย่อมแสดงว่าหน่วยตัวอย่างนั้นค่อนข้างสำคัญและมีผลกระทบต่อการประมาณค่ามากกว่าหน่วยตัวอย่างที่มีโอกาสได้รับการคัดเลือกที่น้อยกว่า 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

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

ν 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

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

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

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

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

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

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

$$=\mu$$

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

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

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

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

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

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

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

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

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

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

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

ดังนั้น

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

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

สำหรับ $i \neq j$

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

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

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

UTTARADIT RAJABHAT UNIVERSITY

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

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

ดังนั้น

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

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

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

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

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

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

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

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

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

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 $\hat{R}=\frac{\bar{y}}{\bar{x}}$ 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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

$$= \mu$$

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 $Z_i = \frac{y_i}{\pi_i}$ และ $Z_j = \frac{y_j}{\pi_j}$ 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

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

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

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

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

เนื่องจาก $E(\hat{R}_{HT}) \neq R$ 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

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

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

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

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

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

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

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

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

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

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

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

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

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

จะได้ว่า

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

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

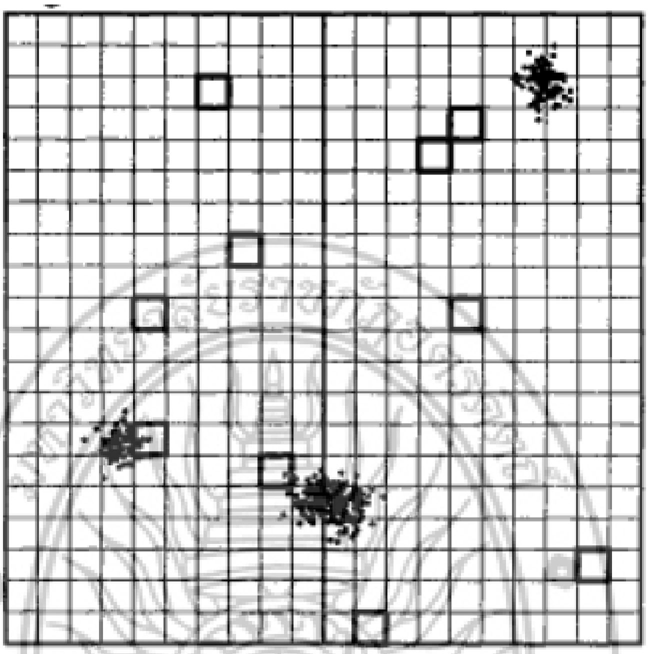
ในทางปฏิบัติเนื่องจากจะไม่ทราบ R ดังนั้นจะประมาณด้วย $\hat{R}$ และ y_i' จะประมาณด้วย $y_i - \hat{R}_{HT} x_i$

ทฤษฎีการสุ่มตัวอย่างแบบกลุ่มปรับเมื่อสุ่มหน่วยตัวอย่างอย่างง่าย

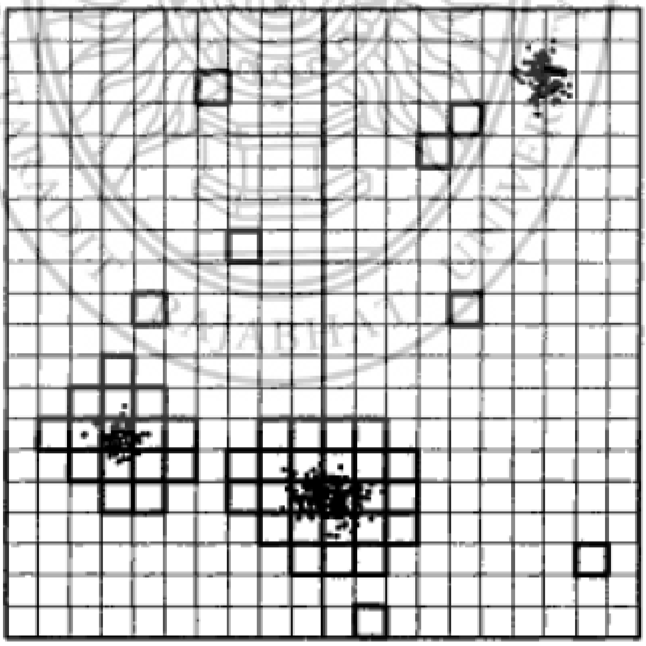
การเลือกหน่วยตัวอย่างในแผนแบบที่ยังไม่ปรับจะกระทำเสร็จก่อนการเก็บรวบรวมข้อมูลจากหน่วยตัวอย่างเหล่านั้น แต่สำหรับแผนการสุ่มตัวอย่างแบบกลุ่มปรับ มีหลักการว่า ลักษณะของข้อมูลที่เก็บรวบรวมได้ น่าจะมีส่วนช่วยทำให้ตัวประมาณมีคุณภาพดีขึ้นโดยการปรับกระบวนการเลือกตัวอย่างให้สามารถนำลักษณะข้อมูลที่เก็บรวบรวมได้มากำหนดหน่วยตัวอย่างที่เก็บรวบรวมเพิ่มขึ้นทั้งในเชิงจำนวน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

แผนการสุ่มตัวอย่างแบบกลุ่มปรับเมื่อสุ่มหน่วยตัวอย่างอย่างง่ายเมื่อแบ่งพื้นที่ออกเป็นชั้นภูมิ คือแผนการสุ่มตัวอย่างที่ปรับกระบวนการสุ่มตัวอย่าง ให้สามารถนำลักษณะของข้อมูลที่เก็บได้มากำหนดหน่วยตัวอย่าง โดยการสุ่มหน่วยตัวอย่างจะใช้วิธีการสุ่มแบบง่าย เช่น กำหนดพื้นที่ที่ต้องการศึกษา และใช้วิธีการสุ่มแบบง่ายในการสำรวจสัตว์ หากพื้นที่ใดพบสัตว์เป็นไปตามเกณฑ์หรือเงื่อนไขที่กำหนด หน่วยทั้งหมดภายในส่วนที่ใกล้เคียงจะถูกเพิ่มเข้าเป็นตัวอย่าง และเป็นค่าสังเกตโดยจะขยายลักษณะเช่นนี้ไปเรื่อย ๆ จนกว่าจะไม่พบสิ่งที่สนใจหรือไม่เป็นไปตามเงื่อนไขที่กำหนด และบางครั้งการขยายการสุ่มหน่วยตัวอย่างในแต่ละกลุ่ม อาจทำให้เกิดการรวมเป็นเครือข่ายเดียวกัน ซึ่งลักษณะการเก็บหน่วยตัวอย่างทั้งหมดที่ได้จากการสุ่มตัวอย่างภายใต้แผนแบบของการเลือกหน่วยตัวอย่างที่ประกอบด้วยการรวมส่วนที่ใกล้เคียงหลาย ๆ ส่วนจะเรียกว่า **กลุ่ม** โดยภายในกลุ่มเป็นการเก็บรวบรวมตัวอย่างย่อยของหน่วยเรียกว่า **เครือข่าย** สำหรับหน่วยตัวอย่างที่สุ่มได้แล้วไม่พบค่าสังเกตจะให้เครือข่ายนั้นมีขนาดเป็น 1 และถ้าหน่วยใดไม่เป็นไปตามเงื่อนไข แต่อยู่ในส่วนใกล้เคียงของหน่วยที่สอดคล้องตามเงื่อนไขจะเรียกหน่วยนั้นว่า **หน่วยขอบ** ความคิดพื้นฐานของแผนการสุ่มตัวอย่างแบบแบ่งชั้นภูมิสามารถอธิบายได้ดังภาพที่ 2 ที่เป็นการประมาณค่าเฉลี่ยของจุดที่สุ่ม โดยแบ่งพื้นที่ออกเป็น 2 ชั้นภูมิ ซึ่งมีขนาด 400 พื้นที่ย่อย และทำการสุ่มหน่วยตัวอย่างของแต่ละชั้นภูมิขนาด 5 หน่วย ด้วยวิธีการเลือกตัวอย่างแบบง่าย ซึ่งแสดงให้เห็นในภาพที่ 2(a) เมื่อสังเกตพบสิ่งที่สนใจจำนวนหนึ่งหรือมากกว่าในหน่วยที่ถูกเลือก หน่วยใกล้เคียงที่ถัดไปทางด้านบน ล่าง ขวา และซ้าย ถูก

เพิ่มเป็นตัวอย่าง เมื่อกระบวนการเสร็จสิ้นสมบูรณ์ แสดงได้ด้วยภาพ 2 (b) (Thompson, 1991, pp. 389-397)



(a)



(b)

ภาพ 2 แสดงแผนการสุ่มตัวอย่างแบบกลุ่มปรับ ในการประมาณค่าของจำนวนจุดสังเกต ในบริเวณที่ศึกษา 400 พื้นที่ย่อย

a) การสุ่มตัวอย่างชั้นภูมิละ 5 หน่วย

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

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

สำหรับงานวิจัยครั้งนี้ทำการศึกษาโดยใช้ตัวประมาณค่าเฉลี่ยที่ดัดแปลงมาจากตัวประมาณ Horvitz–Thompson ซึ่งเป็นตัวประมาณที่ใช้ความน่าจะเป็นไม่เท่ากันแบบไม่แทนที่ในรูปแบบของวิธีประมาณค่าเฉลี่ยแบบง่าย วิธีประมาณค่าเฉลี่ยแบบอัตราส่วน และวิธีประมาณค่าเฉลี่ยแบบอัตราส่วนที่ปรับปรุงโดยวิธีการของ Rao-Blackwell ซึ่งก่อนที่จะกล่าวถึงรายละเอียด และวิธีประมาณค่าดังกล่าวควรทำความเข้าใจเกี่ยวกับสัญลักษณ์ที่ใช้ในการประมาณค่าก่อนเพื่อไม่ให้เกิดความสับสนเกี่ยวกับการใช้สัญลักษณ์

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

1. ตัวประมาณค่าเฉลี่ยแบบง่ายที่ดัดแปลงมาจากตัวประมาณ Horvitz–Thompson

Thompson (1991) ได้ศึกษาวิธีการประมาณค่าเฉลี่ย และความแปรปรวนของตัวประมาณค่าเฉลี่ยแบบง่ายภายใต้แผนการสุ่มตัวอย่างแบบกลุ่มปรับโดยมีรายละเอียดดังทฤษฎี 6

ทฤษฎี 6 การประมาณค่าเฉลี่ยแบบง่ายที่ดัดแปลงมาจากตัวประมาณ Horvitz–Thompson

1. ตัวประมาณที่ไม่เอนเอียงของค่าเฉลี่ยตัวอย่าง

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

เมื่อ $$\alpha_k = 1 - \prod_{k=1}^{L}\binom{N_k - m_{ki}}{n_k}\bigg/\binom{N_k}{n_k}$$

โดยที่

α_k ความน่าจะเป็นที่หน่วยตัวอย่างที่ i อยู่ในเครือข่ายที่ k

u_{y_k} เป็นผลรวมของสิ่งที่สนใจ (y) ในเครือข่ายที่ k

N_k เป็นจำนวนของพื้นที่ย่อยทั้งหมดในชั้นภูมิที่ h

n_k เป็นจำนวนพื้นที่ย่อยตัวอย่าง ในชั้นภูมิที่ h

m_{ki} เป็นจำนวนของพื้นที่ย่อยในเครือข่ายที่ k ชั้นภูมิที่ h ลำดับที่ i

v เป็นจำนวนเครือข่ายที่แตกต่างกันในตัวอย่าง

L เป็นจำนวนชั้นภูมิ

2. ความแปรปรวนของค่าประมาณค่าเฉลี่ยตัวอย่าง

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

เมื่อ $\alpha_{kh} = 1 - \prod_{k=1}^{L}\left\{\binom{N_k - m_{ki}}{n_k} + \binom{N_k - m_{hj}}{n_k} - \binom{N_k - m_{ki} - m_{hj}}{n_k}\right\} \Big/ \binom{N_k}{n_k}$

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

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

พิสูจน์

1. จากแผนการสุ่มตัวอย่างแบบพื้นฐานด้วยความน่าจะเป็นไม่เท่ากันจะทำการสุ่มหน่วยตัวอย่าง n หน่วยจากประชากร N หน่วย และจะนำค่าที่สนใจศึกษา (y_i) ที่ได้จากการสุ่มหน่วยตัวอย่างมาใช้ในการคำนวณ โดยมีค่าความน่าจะเป็น π_i เข้ามาเกี่ยวข้อง แต่สำหรับแผนการสุ่มตัวอย่างแบบกลุ่มปรับจะทำการสุ่มพื้นที่ย่อยจากพื้นที่ทั้งหมดซึ่งภายในพื้นที่ย่อยอาจมีหลายค่าสังเกตหรือหลายหน่วยตัวอย่าง ดังนั้นจึงปรับสัญลักษณ์ เป็นเครือข่ายที่ k และจะนำค่าที่ได้จากการสุ่มพื้นที่ย่อยที่เป็นไปตามเงื่อนไข (u_{y_k}) มาคำนวณโดยที่เครือข่ายที่นำมาคำนวณต้องไม่ซ้ำกัน (v) และสัญลักษณ์ π_i จะเปลี่ยนเป็น α_k ดังนี้

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

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

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

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 $Z_k = \dfrac{u_{y_k}}{\alpha_k}$ และ $Z_h = \dfrac{u_{y_h}}{\alpha_h}$ 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

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

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

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

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

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

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

$$\alpha_k = 1 - \prod_{k=1}^{L}\binom{N_k - m_{ki}}{n_k}\Big/\binom{N_k}{n_k}$$

ในทำนองเดียวกัน

$$\alpha_h = 1 - \prod_{k=1}^{L}\binom{N_k - m_{hj}}{n_k}\Big/\binom{N_k}{n_k}$$

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 $P(A\cup B)=1-\binom{N_k-m_{ki}-m_{hj}}{n_k}\Big/\binom{N_k}{n_k}$

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

$$=\alpha_k+\alpha_h-\left(1-\binom{N_k-m_{ki}-m_{hj}}{n_k}\Big/\binom{N_k}{n_k}\right)$$

$$=\alpha_k+\alpha_h+\binom{N_k-m_{ki}-m_{hj}}{n_k}\Big/\binom{N_k}{n_k}-1$$

$$=1-\frac{\binom{N_k-m_{ki}}{n_k}}{\binom{N_k}{n_k}}+1-\frac{\binom{N_k-m_{hj}}{n_k}}{\binom{N_k}{n_k}}+\frac{\binom{N_k-m_{ki}-m_{hj}}{n_k}}{\binom{N_k}{n_k}}-1$$

$$=1-\frac{\binom{N_k-m_{ki}}{n_k}+\binom{N_k-m_{hj}}{n_k}-\binom{N_k-m_{ki}-m_{hj}}{n_k}}{\binom{N_k}{n_k}}$$

หรือ

$$\alpha_{kh}=P(A\cap B)=1-\left\{\binom{N_k-m_{ki}}{n_k}+\binom{N_k-m_{hj}}{n_k}-\binom{N_k-m_{ki}-m_{hj}}{n_k}\right\}\Big/\binom{N_k}{n_k}$$

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

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

2. ความแปรปรวนของตัวประมาณค่าเฉลี่ยแบบอัตราส่วนเท่ากับ

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

เมื่อ

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

3. ค่าประมาณความแปรปรวนของตัวประมาณค่าเฉลี่ยอัตราส่วนเท่ากับ

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

เมื่อ

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

พิสูจน์

1. กำหนดให้ $\hat{R}_{ACS} = \dfrac{\sum_{k=1}^{\nu}\frac{u_{y_k}}{\alpha_k}}{\sum_{k=1}^{\nu}\frac{u_{x_k}}{\alpha_k}} = \dfrac{\hat{\tau}_y}{\hat{\tau}_x} = \dfrac{\bar{y}^*_{HT}}{\bar{x}^*_{HT}}$ ซึ่งจะเห็นได้ว่า $\hat{R}_{ACS}$ คือค่าเฉลี่ย $\bar{y}^*_{HT}$ หารด้วย $\bar{x}^*_{HT}$ 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

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

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

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

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

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

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

โดยที่

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

ดังนั้น

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

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

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

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

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

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

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

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

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

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

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

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

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

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 ซึ่งการประมาณค่าที่มีการใช้ข้อมูลที่ครบถ้วนนี้จะทำให้เกิดเป็นสถิติพอเพียง 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

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

และ

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

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

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

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

โดยที่

$\bar{y}_e$ เป็นค่าเฉลี่ยของค่า y สำหรับหน่วยขอบตัวอย่าง ในชั้นภูมิที่ h

e_s เป็นจำนวนหน่วยขอบตัวอย่างใน s ในชั้นภูมิที่ h

w_{y_k} เป็นค่าเฉลี่ยของค่า y ในเครือข่ายที่ k ชั้นภูมิที่ h

2. ความแปรปรวนของตัวประมาณค่าเฉลี่ยแบบอัตราส่วนที่ปรับปรุงโดยวิธีการ Rao-Blackwell

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

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

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

$\bar{x}_e$ เป็นค่าเฉลี่ยของค่า x สำหรับหน่วยขอบตัวอย่าง

3. ค่าประมาณความแปรปรวนของตัวประมาณค่าเฉลี่ยแบบอัตราส่วนที่ปรับปรุงโดยวิธีการ Rao-Blackwell

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

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

หรือ

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

พิสูจน์

1. การพิสูจน์จากตัวประมาณค่าเฉลี่ยแบบอัตราส่วนเป็นตัวประมาณค่าเฉลี่ยแบบอัตราส่วนที่ปรับปรุงโดยวิธีการ Rao-Blackwell สามารถกระทำได้ดังนี้

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

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

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

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

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

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

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

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

กำหนดให้ $\hat{R}^{+}_{ACS} = \dfrac{\sum_{k=1}^{v} \frac{m_{ki}.w^{+}_{y_k}}{\alpha_k}}{\sum_{k=1}^{v} \frac{m_{ki}.w^{+}_{x_k}}{\alpha_k}} = \dfrac{\hat{\tau}_y}{\hat{\tau}_x} = \dfrac{\overline{y}^{+}_{HT}}{\overline{x}^{+}_{HT}}$ 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 ดังนั้น โดยปกติ $\hat{R}^{+}_{ACS}$ จะเป็นตัวประมาณที่เอนเอียงสำหรับ R เนื่องจาก

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

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 $\bar{y}^{+}_{R.HT} = \frac{\bar{y}^{+}_{HT}}{\bar{x}^{+}_{HT}}\mu_x = \hat{R}^{+}_{ACS}\mu_x = \bar{y}^{+}_{HT} + \hat{R}_{ACS}(\mu_x - \bar{x}^{+}_{HT})$ (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

$$\begin{aligned}\bar{y}^{+}_{R.HT} - \mu &\approx \bar{y}^{+}_{HT} + R_{ACS}(\mu_x - \bar{x}^{+}_{HT}) - \mu \\ &= \bar{y}^{+}_{HT} + R_{ACS}\mu_x - R_{ACS}\bar{x}^{+}_{HT} - \mu \\ &= \bar{y}^{+}_{HT} + \mu - R_{ACS}\bar{x}^{+}_{HT} - \mu \quad ; R_{ACS}\mu_x = \mu \\ &= \bar{y}^{+}_{HT} - R_{ACS}\ \bar{x}^{+}_{HT}\end{aligned}$$

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

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

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

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

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

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

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

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

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

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

เนื่องจาก

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

ดังนั้น

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

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

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

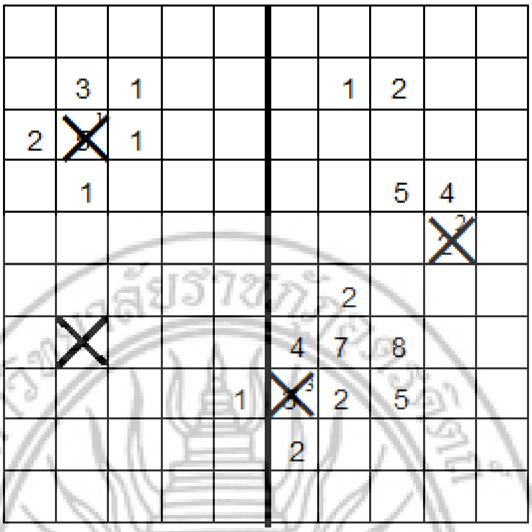
4. ตัวอย่างการคำนวณ

ตัวอย่างที่ 1 ต้องการประมาณจำนวนประชากรกวางผาในป่าแห่งหนึ่งทางตอนเหนือของประเทศไทย โดยพบว่ากวางผามักจะอาศัยอยู่บริเวณที่เป็นเนินเขาหรือผาที่เต็มไปด้วยหิน ดังนั้นจึงกำหนดพื้นที่ที่จะทำการศึกษาที่เป็นเนินเขาหรือผาออกเป็น 100 ตร.หน่วย จากนั้นแบ่งพื้นที่ดังกล่าวออกเป็น 2 ชั้นภูมิ ซึ่งในป่าดังกล่าวมีจำนวนกวางผาแสดงดังภาพ 3

	3	1				1	2		
2	5	1							
	1						5	4	
								2	
						2			
					4	7	8		
				1	3	2	5		
					2				

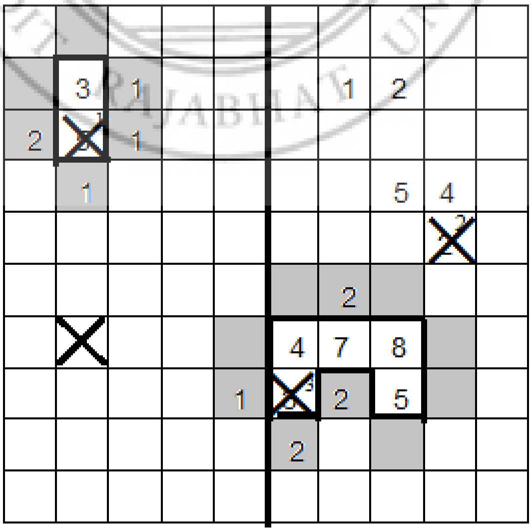
ภาพ 3 แสดงที่ตั้ง และจำนวนของประชากรกวางผากรณีหน่วยขอบแยกกัน

จากนั้นทำการสุ่มตัวอย่างด้วยวิธีการเลือกตัวอย่างแบบง่าย ขนาด 2 พื้นที่ย่อยในแต่ละชั้นภูมิ โดยสุ่มพื้นที่ได้ตรงตำแหน่งดังภาพ 4



ภาพ 4 แสดงตำแหน่งของการสุ่มพื้นที่ย่อยขนาด 2 หน่วยในแต่ละชั้นภูมิด้วยการสุ่มตัวอย่างแบบง่าย

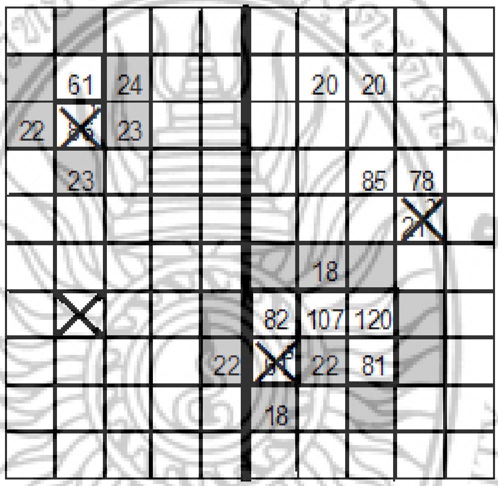
จากนั้นเมื่อทำการสุ่มพื้นที่ย่อยแล้วก็จะพิจารณาพื้นที่ย่อยบริเวณใกล้เคียง โดยที่เงื่อนไขของการรวมเป็นเครือข่ายคือ $y_i \geq 3$ โดยถ้าเป็นไปตามเงื่อนไขก็จะถูกรวมเป็นเครือข่าย แต่ถ้าไม่เป็นไปตามเงื่อนไขก็จะไม่พิจารณา และส่วนที่อยู่รอบเครือข่ายจะเรียกว่าหน่วยขอบ สามารถอธิบายได้ดังภาพ 5



ภาพ 5 แสดงการขยายพื้นที่ในลักษณะ ขึ้น-ลง-ซ้าย-ขวา จนกระทั่งไม่เป็นไปตามเงื่อนไข

จากภาพ 5 พบว่ามีพื้นที่ 3 หน่วยที่พบกวางผาได้แก่พื้นที่ที่ 1, 2 และ 3 ส่วนพื้นที่ที่ 4 ไม่พบกวางผา และเมื่อทำการพิจารณาตามเงื่อนไขถ้าพบกวางผามากกว่าหรือเท่ากับ 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

$N = 100$, $N_1 = N_2 = 50$, $n_1 = 2$, $n_2 = 2$, $\nu = 3$ (เนื่องจากไม่มีเครือข่ายที่ซ้ำกัน) $m_1 = 2$, $m_2 = m_4 = 1$, $m_3 = 5$

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

$$\alpha_k = 1 - \prod_{k=1}^{L} \binom{N_k - m_{ki}}{n_k} \Big/ \binom{N_k}{n_k}$$

$$\alpha_1 = 1 - \binom{50-2}{2} \Big/ \binom{50}{2} = 0.079$$

$$\alpha_2 = \alpha_4 = 1 - \binom{50-1}{2} \Big/ \binom{50}{2} = 0.04$$

$$\alpha_3 = 1 - \binom{50-5}{2} \Big/ \binom{50}{2} = 0.192$$

คำนวณค่าความน่าจะเป็นร่วมของเครือข่ายที่ได้จากการพิจารณาตามเงื่อนไขโดยที่ α_{kh} มีค่าเท่ากับ

$$\alpha_{kh} = 1 - \prod_{k=1}^{L} \left\{ \binom{N_k - m_{ki}}{n_k} + \binom{N_k - m_{hj}}{n_k} - \binom{N_k - m_{ki} - m_{hj}}{n_k} \right\} \Big/ \binom{N_k}{n_k}$$

$$\alpha_{12} = 1 - \left\{ \binom{50-2}{2} + \binom{50-1}{2} - \binom{50-2-1}{2} \right\} \Big/ \binom{50}{2} = 0.0016$$

$$\alpha_{13} = 1 - \left\{ \binom{50-2}{2} + \binom{50-5}{2} - \binom{50-2-5}{2} \right\} \Big/ \binom{50}{2} = 0.0082$$

$$\alpha_{23} = 1 - \left\{ \binom{50-1}{2} + \binom{50-5}{2} - \binom{50-1-5}{2} \right\} \Big/ \binom{50}{2} = 0.0041$$

ดังนั้น

$$\bar{y}^*_{HT} = \frac{1}{N}\sum_{k=1}^{\nu}\frac{u_{y_k}}{\alpha_k} = \frac{1}{100}\sum_{k=1}^{3}\frac{u_{y_k}}{\alpha_k} = \frac{1}{100}\left[\left(\frac{8}{0.079}\right)+\left(\frac{2}{0.04}\right)+\left(\frac{27}{0.192}\right)\right] = 2.92$$

วิธีการประมาณค่าแบบอัตราส่วน

การประมาณจำนวนประชากรกวางผาโดยเฉลี่ยด้วยวิธีการประมาณค่าแบบอัตราส่วนมีค่าเท่ากับ

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

และ

$$\bar{y}^*_{HT} = \frac{1}{N}\sum_{k=1}^{\nu}\frac{u_{y_k}}{\alpha_k} = \frac{1}{100}\sum_{k=1}^{3}\frac{u_{y_k}}{\alpha_k} = \frac{1}{100}\left[\left(\frac{8}{0.079}\right)+\left(\frac{2}{0.04}\right)+\left(\frac{27}{0.192}\right)\right] = 2.92$$

$$\bar{x}^*_{HT} = \frac{1}{N}\sum_{k=1}^{\nu}\frac{u_{x_k}}{\alpha_k} = \frac{1}{100}\sum_{k=1}^{3}\frac{u_{x_k}}{\alpha_k} = \frac{1}{100}\left[\left(\frac{147}{0.079}\right)+\left(\frac{21}{0.04}\right)+\left(\frac{451}{0.192}\right)\right] = 47.35$$

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

ดังนั้น

$$\bar{y}_{R.HT} = \frac{\bar{y}^*_{HT}}{\bar{x}^*_{HT}}\mu_x = \frac{2.92}{47.35}9.94 = 0.613$$

วิธีการประมาณค่าแบบอัตราส่วนที่ปรับปรุงโดยวิธีการ Rao-Blackwell

การประมาณจำนวนประชากรกวางผาโดยเฉลี่ยด้วยวิธีการประมาณค่าแบบอัตราส่วนมีค่าเท่ากับ

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

โดยที่

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

เมื่อ

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

และ

$$\bar{y}_{HT}^{+} = \frac{1}{N}\sum_{k=1}^{\nu}\frac{m_{ki}.w_{y_k}^{+}}{\alpha_k} = \frac{1}{100}\sum_{k=1}^{3}\frac{m_{ki}.w_{y_k}^{+}}{\alpha_k} = \frac{1}{100}\left[\left(\frac{2\times\frac{8}{2}}{0.079}\right)+\left(\frac{1\times\frac{2}{6}}{0.04}\right)+\left(\frac{5\times\frac{27}{5}}{0.192}\right)\right] = 2.50$$

$$\bar{x}_{HT}^{+} = \frac{1}{N}\sum_{k=1}^{\nu}\frac{m_{k}.w_{x_k}^{+}}{\alpha_k} = \frac{1}{100}\sum_{k=1}^{3}\frac{m_{k}.w_{x_k}^{+}}{\alpha_k} = \frac{1}{100}\left[\left(\frac{2\times\frac{147}{2}}{0.079}\right)+\left(\frac{1\times\frac{21}{6}}{0.04}\right)+\left(\frac{5\times\frac{451}{5}}{0.192}\right)\right] = 42.972$$

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

ดังนั้น

$$\bar{y}_{R.HT}^{+} = \frac{2.50}{42.972}9.94 = 0.5783$$

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 รองลงมาคือแบบมีชั้นภูมิ และแบบง่ายตามลำดับ

วิชาญู 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 ผลการวิจัยพบว่า เมื่อเปรียบเทียบประสิทธิภาพในแง่ของความแม่นยำของตัวประมาณค่าเฉลี่ยที่ทำการสุ่มตัวอย่างขั้นต้นอย่างง่ายโดยไม่ใส่คืนภายใต้แผนการสุ่มตัวอย่างแบบกลุ่มปรับกับตัวประมาณค่าเฉลี่ยจากแผนการสุ่มตัวอย่างแบบที่ไม่ปรับ พบว่าตัวประมาณค่าเฉลี่ยแบบกลุ่มปรับมีประสิทธิภาพในแง่ความแม่นยำมากกว่าตัวประมาณค่าเฉลี่ยจากแผนการสุ่มตัวอย่างแบบที่ไม่ปรับทั้ง 3 กรณี และทุก ๆ ขนาดตัวอย่างที่เพิ่มขึ้นจะทำให้ตัวประมาณค่าเฉลี่ยของทั้ง 2 แบบมีความแม่นยำมากขึ้น และการสุ่มตัวอย่างแบบกลุ่มปรับเหมาะสมกับการสุ่มตัวอย่างของสิ่งตัวอย่างที่เราสนใจที่อยู่เกาะกลุ่มกัน ซึ่งเป็นลักษณะของสิ่งที่หายาก เช่น ฝูงสัตว์ พืชต่าง ๆ ฟอสซิล รวมทั้งแร่ธาตุ เป็นต้น ส่วนการสุ่มตัวอย่างแบบที่ไม่ปรับเหมาะสมกับการสุ่มตัวอย่างของสิ่งที่เราสนใจ ซึ่งมีลักษณะที่อยู่กระจายตัวกัน

Thompson (1990) ทำการศึกษา และเปรียบเทียบประสิทธิภาพของแผนการสุ่มตัวอย่างแบบกลุ่มปรับกับแผนการสุ่มตัวอย่างแบบกลุ่มธรรมดาที่ยังไม่มีการปรับกลุ่ม และเปรียบเทียบประสิทธิภาพตัวประมาณค่าเฉลี่ยที่ดัดแปลงมาจากตัวประมาณ Hansen-Hurwitz และ Horvitz-

Thompson ที่หน่วยตัวอย่างใช้วิธีการสุ่มอย่างง่าย เกณฑ์ที่ใช้ในการเปรียบเทียบคือ ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย จากการวิจัยพบว่า ถ้าใช้ตัวประมาณของแผนการสุ่มตัวอย่างแบบกลุ่มธรรมดาที่ยังไม่มีการปรับกลุ่มนั้นมาประมาณค่าเฉลี่ยสิ่งที่สนใจที่มีลักษณะหายาก และอยู่รวมกันเป็นแล้วนั้น จะทำให้เกิดความเอนเอียง และพบว่าตัวประมาณค่าเฉลี่ยที่ดัดแปลงมาจากตัวประมาณ Horvitz-Thompson มีค่าความแปรปรวนต่ำกว่าตัวประมาณค่าเฉลี่ยที่ดัดแปลงมาจากตัวประมาณ Hansen-Hurwitz

Thompson (1991) ทำการศึกษา และเปรียบเทียบประสิทธิภาพของแผนการสุ่มตัวอย่างแบบกลุ่มปรับเมื่อสุ่มตัวอย่างแบบแบ่งชั้นภูมิ และเปรียบเทียบประสิทธิภาพตัวประมาณค่าเฉลี่ยที่ดัดแปลงมาจากตัวประมาณ Hansen-Hurwitz และ Horvitz-Thompson ที่หน่วยตัวอย่างใช้วิธีการสุ่มอย่างง่าย เกณฑ์ที่ใช้ในการเปรียบเทียบคือ ค่าความแปรปรวน จากการวิจัยพบว่า พบว่าตัวประมาณค่าเฉลี่ยที่ดัดแปลงมาจากตัวประมาณ 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

บทที่ 3
วิธีดำเนินงานวิจัย

การวิจัยครั้งนี้เป็นการเปรียบเทียบประสิทธิภาพวิธีประมาณค่าเฉลี่ยที่ดัดแปลงมาจากตัวประมาณ Horvitz-Thompson ภายใต้แผนการสุ่มตัวอย่างแบบกลุ่มปรับด้วยวิธีการประมาณค่าอย่างง่าย วิธีการประมาณค่าแบบอัตราส่วน และวิธีการประมาณค่าแบบอัตราส่วนที่ปรับปรุงโดยวิธีการ Rao-Blackwell เมื่อสุ่มหน่วยตัวอย่างแบบแบ่งชั้นภูมิ โดยใช้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยเป็นเกณฑ์ในการเปรียบเทียบเพื่อหาวิธีประมาณค่าที่ดีที่สุดในแต่ละสถานการณ์ ข้อมูลที่ใช้ในการวิจัยได้จากการจำลองข้อมูลให้มีลักษณะตามสถานการณ์ที่กำหนดโดยใช้โปรแกรม Minitab 14 และการประมาณค่าพารามิเตอร์ใช้โปรแกรม MATLAB ซึ่งมีการวางแผน และกำหนดขั้นตอนดำเนินการวิจัยดังนี้

1. การสร้างข้อมูลในการวิจัย
2. ขั้นตอนการเลือกตัวอย่าง
3. การประมาณค่าเฉลี่ยภายใต้แผนการสุ่มตัวอย่างแบบกลุ่มปรับ
4. เกณฑ์การตัดสินใจ
5. แผนผังขั้นตอนดำเนินงานวิจัย

โดยมีรายละเอียดดังนี้

การสร้างข้อมูลในการวิจัย

ในการสร้างข้อมูลที่ใช้ในการวิจัยจะแบ่งออกเป็น 2 สถานการณ์หลักดังที่กล่าวไว้ในขอบเขตการวิจัยโดยในแต่ละสถานการณ์สร้างประชากรให้มีลักษณะที่แตกต่างกัน ภายใน 100 พื้นที่ย่อยที่มีจำนวน 2 ชั้นภูมิ

ตัวแปร Y คือ ตัวแปรที่ต้องการศึกษา เป็นข้อมูลเชิงปริมาณ

ตัวแปร X คือ ตัวแปรช่วยของสิ่งที่ต้องการศึกษา เป็นข้อมูลเชิงปริมาณในพื้นที่ย่อย

และในการสร้างข้อมูลเชิงปริมาณของ X และ Y จะต้องพิจารณาเงื่อนไขคือกำหนดให้ระดับความสัมพันธ์ระหว่างข้อมูล X และ Y ซึ่งแบ่งออกเป็น 3 ระดับดังนี้

ความสัมพันธ์ระดับต่ำ	อยู่ในช่วง (0 , 0.4)
ความสัมพันธ์ระดับปานกลาง	อยู่ในช่วง [0.4, 0.7)
ความสัมพันธ์ระดับสูง	อยู่ในช่วง [0.7, 1.00)

ในการตรวจสอบระดับความสัมพันธ์ดังกล่าวจะใช้วิธีการหาค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างสองตัวแปร ถ้าข้อมูลที่จำลองขึ้นมีความสัมพันธ์ไม่ตรงตามที่กำหนดไว้ ก็จะทำการจำลองข้อมูลจนกระทั่งเป็นไปตามข้อตกลงทุกประการ ส่วนประชากรที่ใช้ในการวิจัยมีหลักการสร้างดังนี้

หลักการสร้างประชากรที่มีลักษณะที่หายาก

กระบวนการปัวซงคลัสเตอร์ หรือเรียกอีกอย่างหนึ่งว่า Centre – Satellite Process เป็นกระบวนการจำลองสถานการณ์โดยอาศัยระยะห่างของค่าสังเกตแต่ละตัว มาสร้างสถานการณ์หรือจำลองประชากรให้มีลักษณะเกาะกลุ่มกันภายในพื้นที่ที่กำหนด พัฒนาครั้งแรกโดย Neyman และ Scott (1985) ซึ่งต่อมา A.D.Cliff และ J.K.Ord ได้สรุปกระบวนการปัวซงคลัสเตอร์โดยการจำลองด้วยเครื่องคอมพิวเตอร์ซึ่งทำอย่างต่อเนื่องโดยมีหลักการที่สำคัญ 3 ขั้นตอนคือ

1. หาจำนวนจุดหลัก สร้างด้วยกระบวนการปัวซง ที่มีความหนาแน่น λ ต่อหนึ่งหน่วยพื้นที่ มี 2 ขั้นตอนคือ

1.1 หาตำแหน่งของเหตุการณ์ ซึ่งจำลองจากตัวแปรสุ่มอิสระที่มีฟังก์ชันความหนาแน่นแบบ Exponential (λ) และจำลองจากโปรแกรม Minitab 14 ด้วย $\lambda = 30$ ในช่วงพื้นที่ 100 หน่วย และจำลองประชากรออกเป็น 2 สถานการณ์หลักซึ่งแต่ละสถานการณ์มีลักษณะของสิ่งที่สนใจกระจายอยู่ในรูปแบบที่แตกต่างกัน

1.2 นับจำนวนเหตุการณ์ที่เกิดในช่วง (0,100) โดยเอาตำแหน่งที่ไม่เกิน 100 เท่านั้น

2. หาจำนวนบริวาร ของแต่ละจุดหลักซึ่งเป็นค่าสังเกตที่เป็นอิสระต่อกันมาจากการแจกแจงความน่าจะเป็นที่กำหนด โดยงานวิจัยนี้ใช้การแจกแจงแบบปัวซง และจำลองจากโปรแกรม Minitab 14 ด้วย $\lambda = 30$

3. หาที่ตั้งของบริวาร มีความสัมพันธ์กับตำแหน่งของจุดหลักซึ่งเป็นอิสระต่อกัน มีการแจกแจงแบบปกติ 2 ตัวแปร

ในการดำเนินการสร้างประชากรที่มีลักษณะที่หายาก ผู้วิจัยขอนำเสนอรายละเอียดการสร้างประชากรเฉพาะกรณีหน่วยขอบแยกกัน 2 กลุ่มเท่านั้น ส่วนประชากรกรณีอื่น ๆ สามารถพิจารณาได้ในลักษณะเดียวกัน

ขั้นตอนการจำลองประชากรที่มีลักษณะหายากมีรายละเอียดดังนี้

1. หาจำนวนจุดหลัก สร้างด้วยกระบวนการปัวซง ที่มีความหนาแน่น λ ต่อหนึ่งหน่วยพื้นที่ (Unit Area) มี 2 ขั้นตอนคือ

1.1 หาตำแหน่งของเหตุการณ์ซึ่งจำลองจากตัวแปรสุ่มอิสระที่มีฟังก์ชันความหนาแน่นแบบ Exponential (λ) และจำลองจากโปรแกรม Minitab 14 ด้วย $\lambda = 30$ ได้ผลลัพธ์ดังนี้

MTB > Random 5 c1

SUBC > Exponential 30

MTB > print c1

C1

38.2631 35.0943 39.2180 82.6139 65.6530

1.2 นับจำนวนเหตุการณ์ที่เกิดในช่วง (0,100) โดยเอาตำแหน่งที่ไม่เกิน 100 เท่านั้น ผลลัพธ์ที่ได้คือมีจำนวนจุดหลัก 2 จุด คือ (38.2631 + 35.0943 = 73.3574) ซึ่งไม่รวมจุดที่ 3 ที่มีค่า 39.2180 ซึ่งเกิน 100

2. หาจำนวนบริวาร ของแต่ละจุดหลักซึ่งเป็นค่าสังเกตที่เป็นอิสระต่อกันมาจากการแจกแจงความน่าจะเป็นที่กำหนด โดยงานวิจัยนี้ใช้การแจกแจงแบบปัวซง และจำลองจากโปรแกรม Minitab 14 ด้วย $\lambda = 30$ ได้ผลลัพธ์ดังนี้

MTB > Random 5 c1

SUBC > Exponential 30

MTB > print c1

C1

24 28

ดังนั้น

กลุ่มที่ 1 มีจำนวนบริวาร 24 ตัว

กลุ่มที่ 2 มีจำนวนบริวาร 28 ตัว

3. หาที่ตั้งของบริวาร มีความสัมพันธ์กับตำแหน่งของจุดหลักซึ่งเป็นอิสระต่อกัน มีการแจกแจงแบบปกติ 2 ตัวแปร จำลองจากโปรแกรม Minitab 14 ได้ผลลัพธ์ดังนี้

3.1 Choose **Data > Copy > Columns to Matrix**.

3.2 In **Copy from columns**, enter *C2-C3*.

3.3 In **Store Copied Data,** under **In current worksheet, in matrix**, enter *M1*. Click **OK**.

3.4 Choose **Calc > Random Data > Multivariate Normal**.

3.5 In **Generate _ rows of data**, enter 24 และ 28

3.6 In **Store in column(s):** enter *C7-C8* และ *C10-C11*

3.7 In **Mean column:** enter *C1*.

3.8 In **Variance-Covariance Matrix:** enter *M1*. Click **OK**.

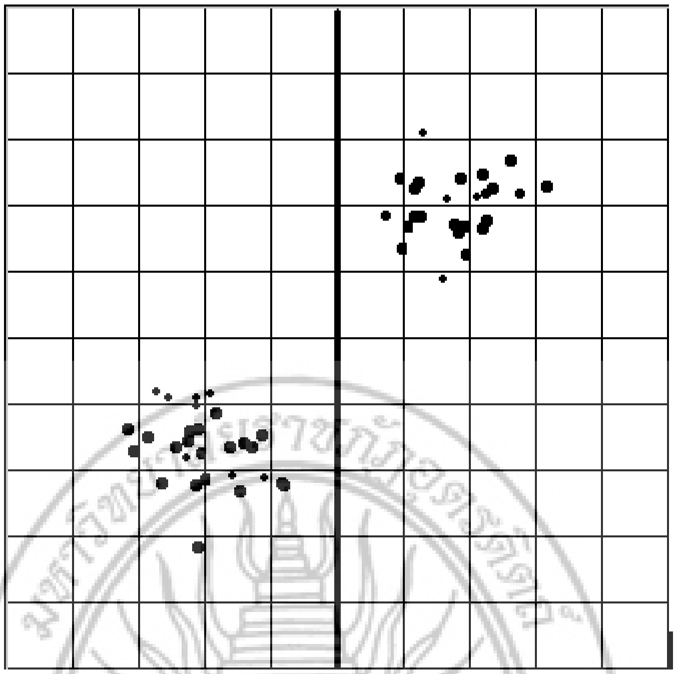
จุดหลักที่ 1 มีจำนวน 24 ตัว มีตำแหน่งดังนี้

ตัวแปร x	ตัวแปร y	ตัวแปร x	ตัวแปร y
-0.13211	-0.54078	-0.6658	-0.4388
-0.55364	0.25124	1.75325	1.6368
0.58845	0.4153	0.44423	-0.07828
-0.84887	-0.46815	0.26691	0.66925
-0.38235	-0.01228	-0.64754	-1.13505
0.27358	0.21053	-1.4776	-1.32582
-0.71009	-0.60903	1.54749	0.95682
0.03001	-0.64639	-0.28705	0.15097
-0.06201	-1.06288	0.07636	-0.04871
-0.89496	-0.70492	0.07769	0.25502
-1.06091	-1.24888	-1.94968	-1.98580
0.92397	1.40132	-0.17506	-0.60014

จุดหลักที่ 2 มีจำนวน 28 ตัว มีตำแหน่งดังนี้

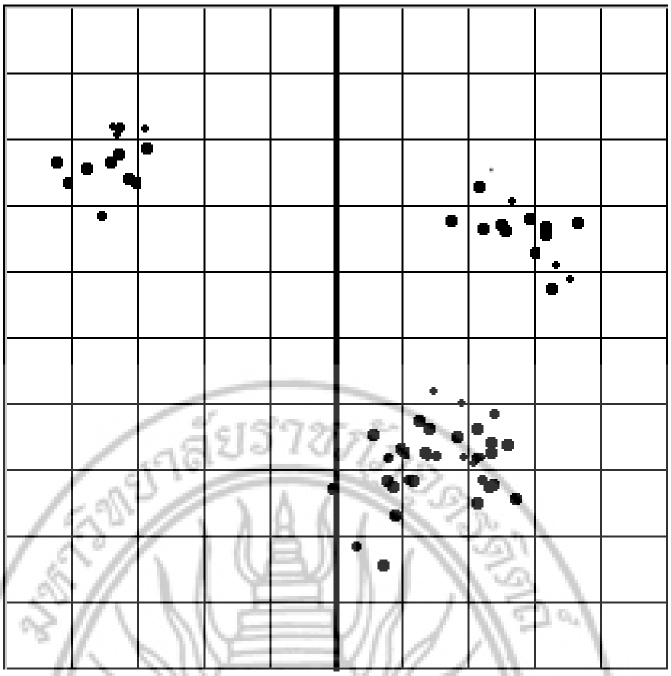
ตัวแปร x	ตัวแปร y	ตัวแปร x	ตัวแปร y
-0.68739	-0.60039	-0.86609	-0.99349
1.52703	-0.02338	-0.07764	-0.95985
0.47187	1.3179	-0.85202	-1.66508
-1.42746	-0.87319	0.83477	1.07267
-0.30206	0.36178	-0.28761	-0.59992
-0.6064	0.33108	0.75047	1.05844
-0.39166	-0.85091	0.90159	0.58561
-2.01651	-1.94199	0.07714	0.50855
0.00719	-0.11800	-0.32718	0.69242
0.73885	1.34408	0.98713	0.13160
-0.88252	-1.34563	-0.29595	0.19970
-0.05057	-0.17976	-0.44148	-1.13529
-0.87833	-1.14517	0.17933	-0.10763
-0.55835	-1.73831	1.62260	1.25489

กรณีประชากรสถานการณ์ที่ 1 หน่วยขอบแยกกัน แสดงดังแผนภาพ 7, 8 และ 9 ตามลำดับ โดยสามารถพิจารณารายละเอียดได้ดังนี้



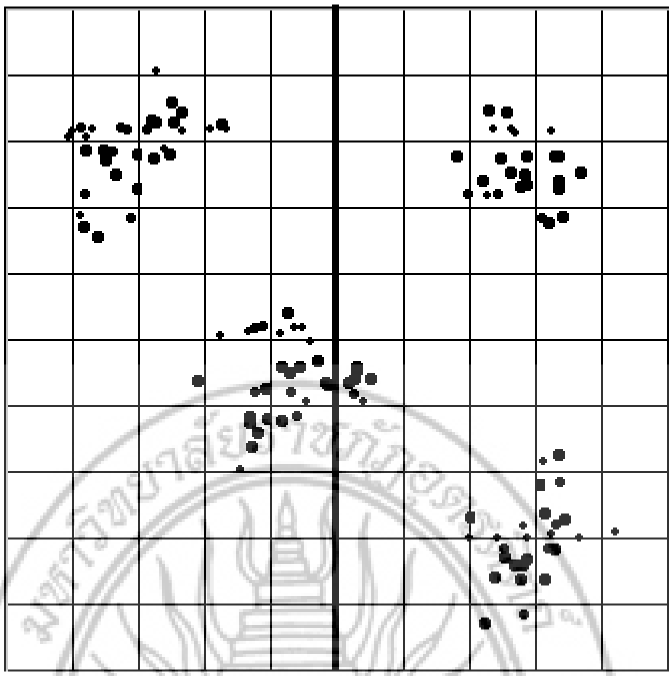
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	1	0	0	0
0	0	0	0	0	1	4	6	1	0
0	0	0	0	0	3	5	2	0	0
0	0	0	0	0	0	1	0	0	0
0	0	4	1	0	0	0	0	0	0
0	2	7	5	0	0	0	0	0	0
0	0	3	3	2	0	0	0	0	0
0	0	1	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0

ภาพ 7 แสดงที่ตั้ง และจำนวนของประชากรสถานการณ์ที่ 1 กรณีหน่วยขอบแยกกัน 2 กลุ่ม



0	0	0	0	0	0	0	0	0	0
0	3	1	0	0	0	0	0	0	0
2	5	1	0	0	0	0	2	0	0
0	1	0	0	0	0	1	5	4	0
0	0	0	0	0	0	0	0	2	0
0	0	0	0	0	0	2	0	0	0
0	0	0	0	0	4	7	8	0	0
0	0	0	0	1	3	2	5	0	0
0	0	0	0	0	2	0	0	0	0
0	0	0	0	0	0	0	0	0	0

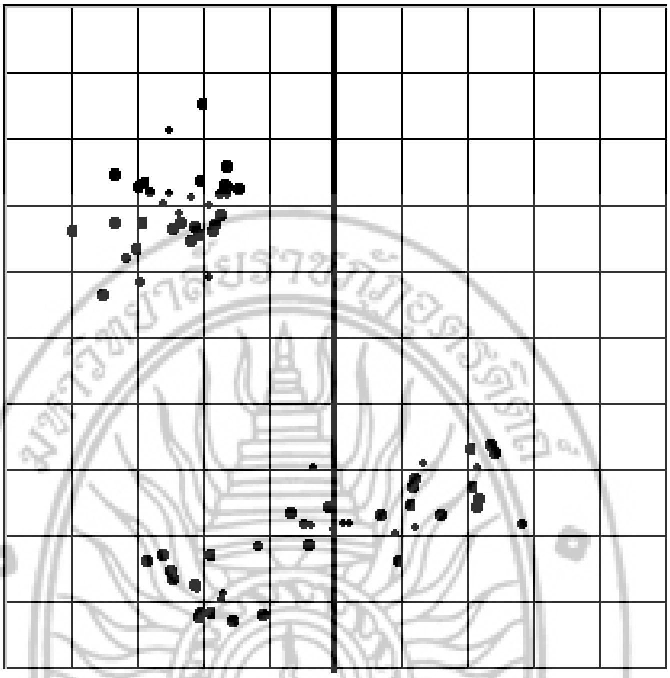
ภาพ 8 แสดงที่ตั้ง และจำนวนของประชากรสถานการณ์ที่ 1 กรณีหน่วยขอบแยกกัน 3 กลุ่ม



0	0	1	0	0	0	0	0	0	0
2	5	7	3	0	0	0	5	1	0
0	8	3	0	0	0	2	9	5	0
0	4	0	0	0	0	0	0	3	0
0	0	0	4	5	0	0	0	0	0
0	0	1	2	8	7	0	0	0	0
0	0	0	6	2	0	0	0	2	0
0	0	0	0	0	0	2	4	6	1
0	0	0	0	0	0	0	7	3	0
0	0	0	0	0	0	0	2	0	0

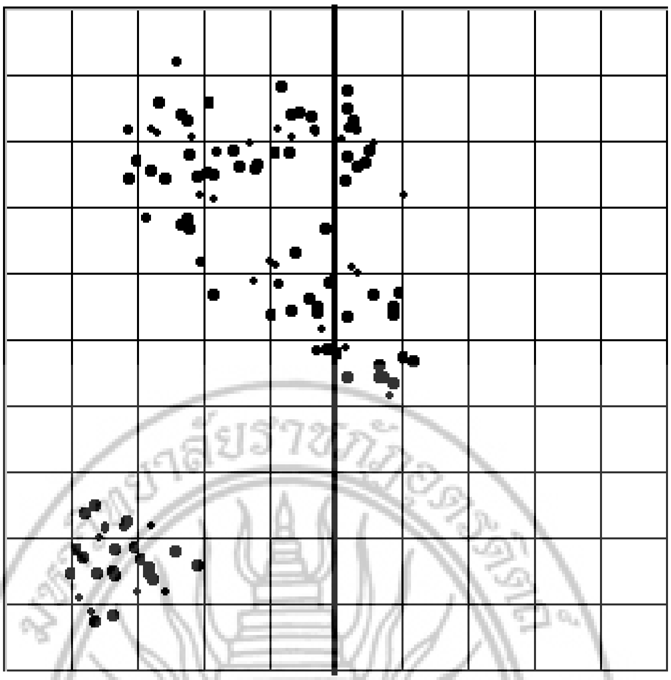
ภาพ 9 แสดงที่ตั้ง และจำนวนของประชากรสถานการณ์ที่ 1 กรณีหน่วยขอบแยกกัน 4 กลุ่ม

กรณีประชากรสถานการณ์ที่ 2 หน่วยขอบร่วมกัน แสดงดังแผนภาพ 10, 11 และ 12 ตามลำดับ โดยสามารถพิจารณารายละเอียดได้ดังนี้



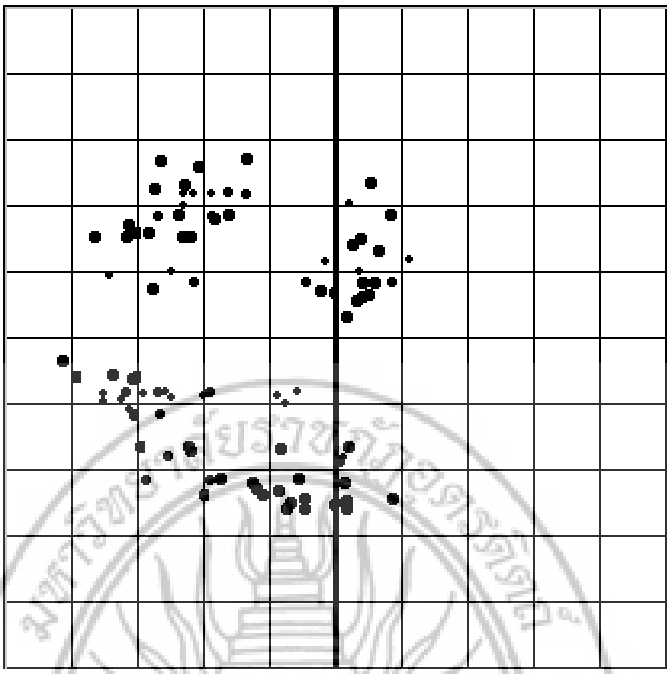
0	0	0	0	0	0	0	0	0	0
0	0	2	0	0	0	0	0	0	0
0	3	5	7	0	0	0	0	0	0
1	4	6	3	0	0	0	0	0	0
0	2	0	1	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	1	0	2	3	0	0
0	0	0	0*	5	4	6	3	0	0
0	0	7	3	1*	1	0	0	0	0
0	0	3	2	0	0	0	0	0	0

ภาพ 10 แสดงที่ตั้ง และจำนวนของประชากรสถานการณ์ที่ 2 กรณีหน่วยขอบร่วมกัน 2 กลุ่ม



0	0	1	0	0	0	0	0	0	0
0	1	7	1*	8	7	0	0	0	0
0	2	6	8	1*	6	0	0	0	0
0	0	5	1*	4	2*	0	0	0	0
0	0	0*	3	7	5	0	0	0	0
0	0	0	0	3	8	1	0	0	0
0	0	0	0	0	0	0	0	0	0
0	7	1	0	0	0	0	0	0	0
2	9	6	0	0	0	0	0	0	0
0	3	0	0	0	0	0	0	0	0

ภาพ 11 แสดงที่ตั้ง และจำนวนของประชากรสถานการณ์ที่ 2 กรณีหน่วยขอบร่วมกัน 3 กลุ่ม



0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	7	4	0	2	0	0	0	0
0	5	6	3	*1	5	1	0	0	0
0	*0	*2	*0	3	7	0	0	0	0
2	7	5	*1	3	*0	0	0	0	0
0	3	4	*0	*1	3	0	0	0	0
0	0	*2	5	7	5	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0

ภาพ 12 แสดงที่ตั้ง และจำนวนของประชากรสถานการณ์ที่ 2 กรณีหน่วยขอบร่วมกัน 4 กลุ่ม

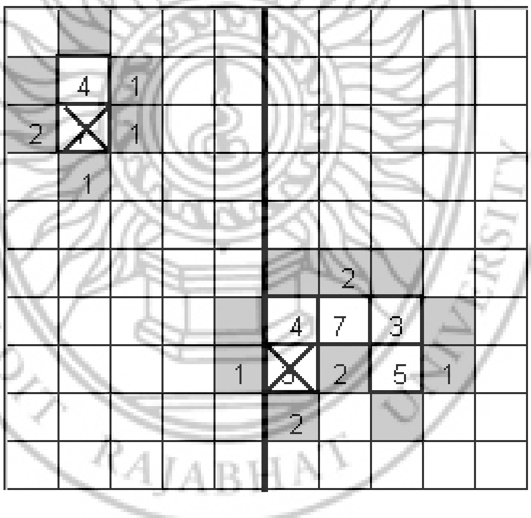
ขั้นตอนการเลือกตัวอย่าง

กำหนดขนาดพื้นที่ย่อยที่ใช้ในการวิจัย คือ 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 16, 20 และ 24 หน่วยในแต่ละชั้นภูมิ และเลือกหน่วยตัวอย่างภายใต้แผนการสุ่มตัวอย่างแบบกลุ่มปรับดังนี้

1. เลือกหน่วยตัวอย่างด้วยวิธีการสุ่มตัวอย่างแบบง่ายไม่แทนที่จากประชากรที่มีลักษณะที่หายาก 2 สถานการณ์หลัก

2. ถ้าพบว่าหน่วยตัวอย่างที่ได้จากการสุ่มเป็นไปตามเงื่อนไขคือ $C = \{y : y \geq 3\}$ โดยที่ให้ C เป็น set ของค่าที่มากกว่าหรือเท่ากับ 3 โดยถ้า y_i ที่อยู่ใน C มีค่ามากกว่าหรือเท่ากับ 3 ที่เป็นไปตามเงื่อนไขที่กำหนดก็จะพิจารณาหน่วยบริเวณใกล้เคียงเข้าไปเป็นหน่วยตัวอย่าง

โดยรูปแบบ และวิธีการเลือกตัวอย่างแสดงรายละเอียดดังภาพ 17 ซึ่งเป็นการสุ่มพื้นที่ย่อยด้วยวิธีการสุ่มตัวอย่างแบบง่าย จากนั้นก็จะพิจารณาพื้นที่ย่อยบริเวณใกล้เคียง โดยที่เงื่อนไขของการรวมเป็นเครือข่ายคือ $y_i \geq 3$ ถ้าเป็นไปตามเงื่อนไขก็จะถูกรวมเป็นเครือข่าย แต่ถ้าไม่เป็นไปตามเงื่อนไขก็จะไม่พิจารณา และส่วนที่อยู่รอบเครือข่ายที่เป็นสีเทาจะเรียกว่าหน่วยขอบ



ภาพ 13 แสดงการเลือกตัวอย่าง และการขยายพื้นที่ตามเงื่อนไขที่กำหนด

การประมาณค่าเฉลี่ยภายใต้แผนการสุ่มตัวอย่างแบบกลุ่มปรับ

1. วิธีประมาณค่าอย่างง่าย เป็นการประมาณค่าเฉลี่ยของตัวแปรที่สนใจศึกษา ที่ตัวประมาณมีการดัดแปลงให้เหมาะสมกับแผนการสุ่มตัวอย่างแบบกลุ่มปรับ

2. วิธีการประมาณค่าแบบอัตราส่วน เป็นวิธีการประมาณค่าเฉลี่ยของตัวแปรที่สนใจศึกษา และมีการใช้ประโยชน์จากตัวแปรช่วยมาพิจารณาในการประมาณค่า

3. วิธีการประมาณค่าแบบอัตราส่วนที่ปรับปรุงโดยวิธีการ Rao-Blackwell เป็นวิธีการประมาณค่าที่ใช้ประโยชน์จากตัวแปรช่วย และใช้วิธีการ Rao-Blackwell มาปรับปรุงวิธีการเพื่อทำให้ค่าประมาณที่ได้เป็นสถิติพอเพียงต่ำสุด และมีความแปรปรวนต่ำ ซึ่งเป็นคุณสมบัติที่ดีทางสถิติ

เกณฑ์การตัดสินใจ

เกณฑ์ที่ใช้ในการประเมินว่าวิธีการประมาณค่าในรูปแบบการสุ่มตัวอย่างแบบกลุ่มปรับวิธีการใดที่มีประสิทธิภาพมากที่สุดจะพิจารณาจากค่าความคลาดเคลื่อนกำลังสองเฉลี่ยโดยวิธีที่ให้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุดจะเป็นวิธีการประมาณค่าที่มีประสิทธิภาพมากที่สุด โดยมีวิธีการคำนวณดังนี้

$$Bias(\bar{y}) = \frac{\sum_{i=1}^{1000} Bias(\bar{y}_t)}{1000} \quad \text{โดยที่} \quad Bias(\bar{y}_t) = \bar{y}_t - \mu$$

เมื่อ

$\bar{y}_t$ คือ ค่าประมาณค่าเฉลี่ยรอบที่ t ของแต่ละวิธี

t คือ จำนวนรอบของการทำซ้ำ

โดยที่

$Bias(\bar{y})$ คือ ค่าประมาณค่าความเอนเอียงของค่าประมาณค่าเฉลี่ย

$Bias(\bar{y}_t)$ คือ ค่าประมาณค่าความเอนเอียงของค่าประมาณค่าเฉลี่ยรอบที่ t

ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย

$$MSE(\bar{y}) = \frac{\sum_{i=1}^{1000} MSE(\bar{y}_t)}{1000} \quad \text{โดยที่} \quad MSE(\bar{y}_t) = Var(\bar{y}_t) + Bias(\bar{y}_t)^2$$

โดยที่

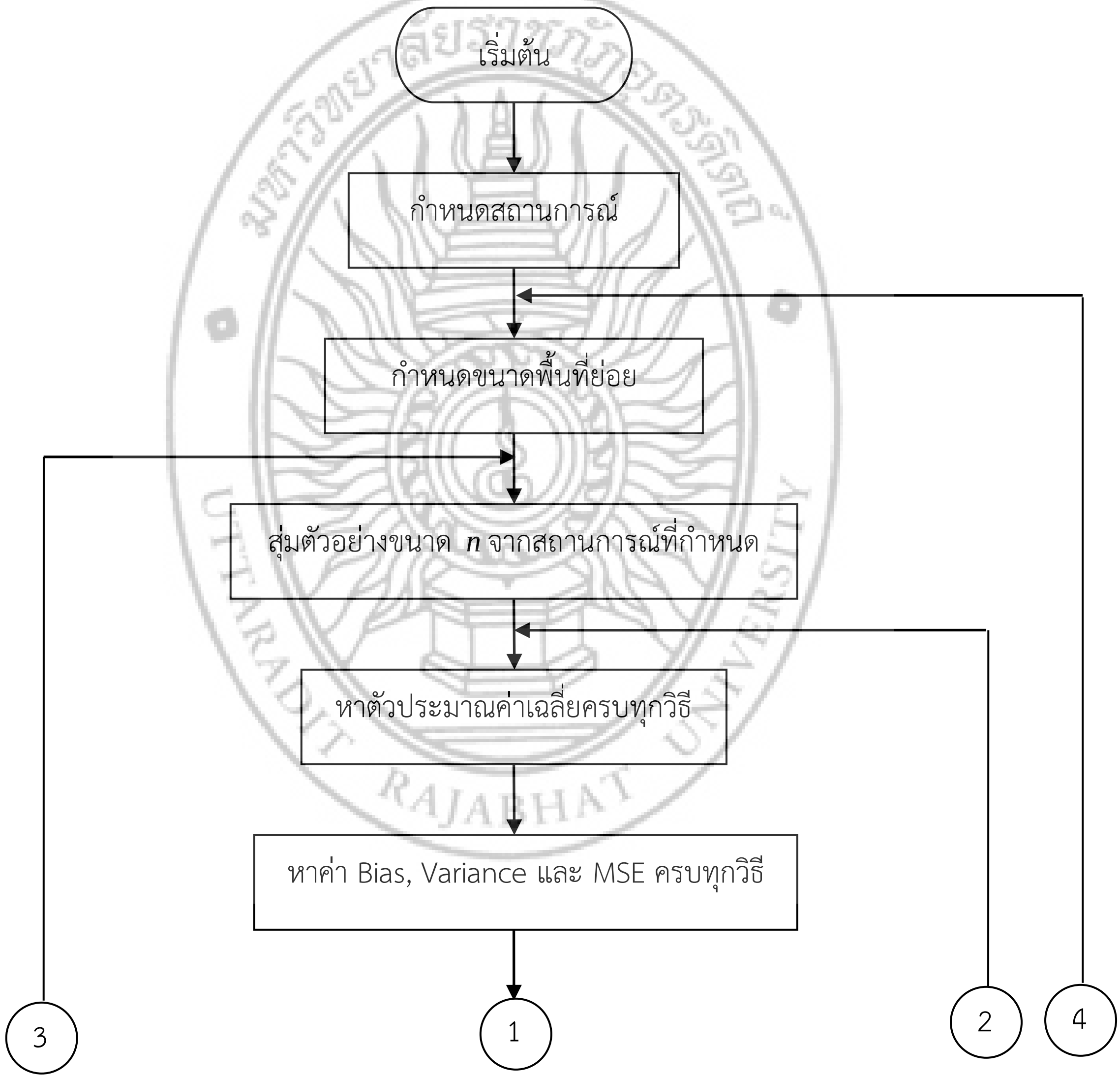
$MSE(\bar{y})$ คือ ค่าประมาณความคลาดเคลื่อนกำลังสองของค่าประมาณค่าเฉลี่ย

$MSE(\bar{y}_t)$ คือ ค่าความคลาดเคลื่อนกำลังสองของค่าเฉลี่ยรอบที่ t

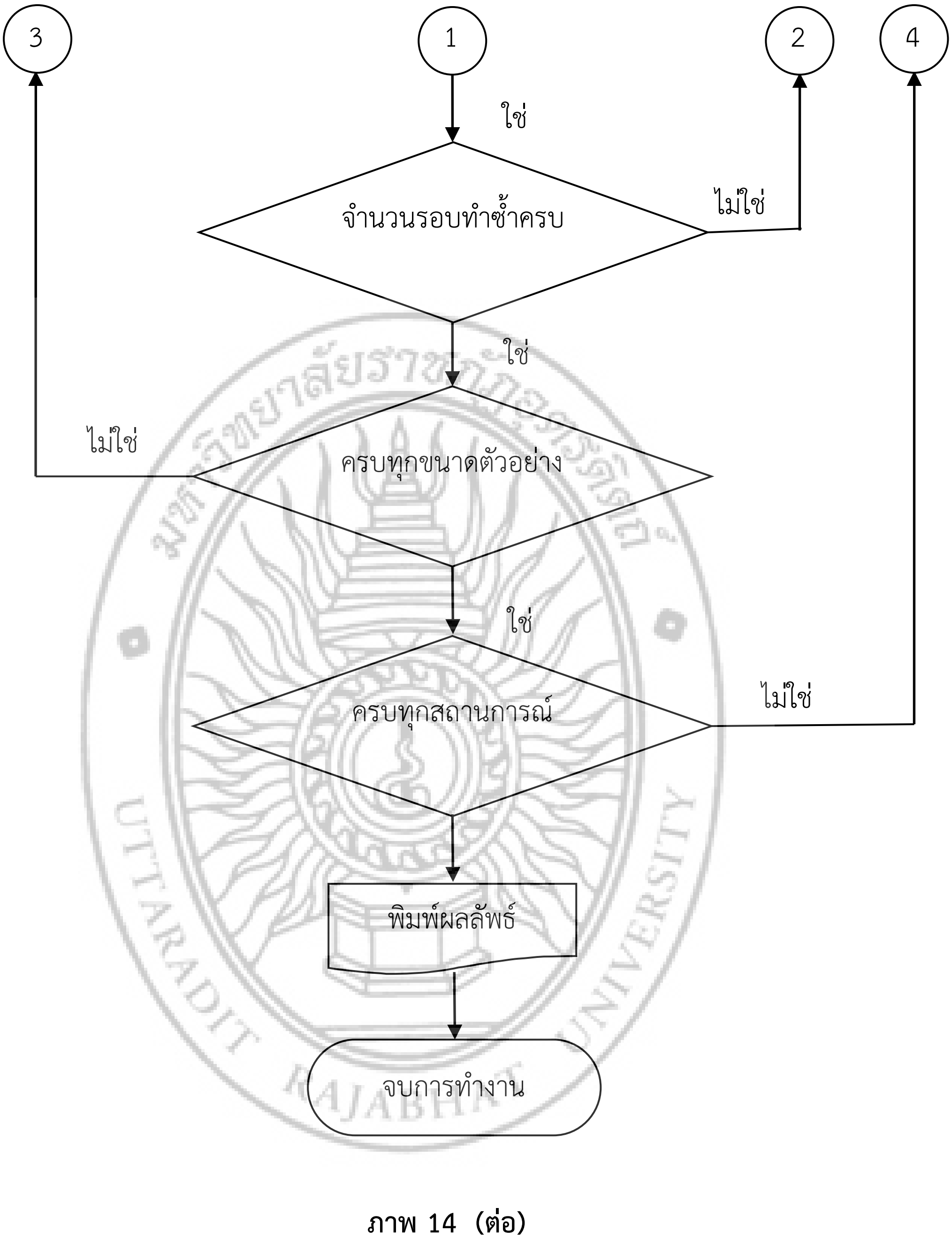
$Var(\bar{y}_t)$ คือ ค่าประมาณความแปรปรวนของค่าประมาณค่าเฉลี่ยรอบที่ t ของแต่ละวิธีที่ได้กล่าวไว้ในทฤษฎีบทที่ 2

แผนผังขั้นตอนดำเนินงานวิจัย

ในการวิจัยครั้งนี้ ผู้วิจัยได้ใช้โปรแกรม MATLAB ในการสร้าง และวิเคราะห์ข้อมูลต่าง ๆ ซึ่งสามารถแสดงแผนผังขั้นตอนดำเนินงานวิจัยได้ดังนี้



ภาพ 14 แสดงแผนผังแสดงลำดับการทำงานหลักของโปรแกรม



ภาพ 14 (ต่อ)

UTTARADIT RAJABHAT UNIVERSITY

บทที่ 4
ผลการวิจัย

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพวิธีประมาณค่าเฉลี่ยที่ดัดแปลงมาจากตัวประมาณ Horvitz-Thompson ภายใต้แผนการสุ่มตัวอย่างแบบกลุ่มปรับด้วย วิธีประมาณค่าอย่างง่าย วิธีประมาณค่าแบบอัตราส่วน และวิธีประมาณค่าแบบอัตราส่วนที่ปรับปรุงโดยวิธีการ Rao-Blackwell เมื่อสุ่มหน่วยตัวอย่างแบบแบ่งชั้นภูมิ โดยใช้ความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) เป็นเกณฑ์ในการเปรียบเทียบ โดยวิธีประมาณค่าที่มีค่าความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุดจะเป็นวิธีประมาณค่าที่มีประสิทธิภาพมากที่สุดในแต่ละสถานการณ์ ในงานวิจัยนี้กำหนดสัญลักษณ์แทนความหมายของค่าต่าง ๆ ดังต่อไปนี้

n แทน ขนาดพื้นที่ย่อย

SE แทน วิธีการประมาณค่าอย่างง่าย

RA แทน วิธีการประมาณค่าแบบอัตราส่วน

RB แทน วิธีการประมาณค่าแบบอัตราส่วนที่ปรับปรุงโดยวิธีการ Rao-Blackwell

โดยผลการวิจัยสามารถแบ่งออกเป็น 3 ตอนดังนี้

ตอนที่ 1 ผลการเปรียบเทียบความคลาดเคลื่อนกำลังสองเฉลี่ย ภายใต้ประชากรสถานการณ์ที่ 1 กรณีหน่วยขอบแยกกัน

ตอนที่ 2 ผลการเปรียบเทียบความคลาดเคลื่อนกำลังสองเฉลี่ย ภายใต้ประชากรสถานการณ์ที่ 2 กรณีหน่วยขอบร่วมกัน

ตอนที่ 3 สรุปจำนวนความถี่วิธีประมาณค่าเฉลี่ย จำแนกตามระดับความสัมพันธ์ จำนวนหน่วยขอบ และสถานการณ์ที่หน่วยขอบแยกกัน และหน่วยขอบร่วมกัน

ตอนที่ 1 ผลการเปรียบเทียบความคลาดเคลื่อนกำลังสองเฉลี่ย ภายใต้ประชากรสถานการณ์ที่ 1 กรณีหน่วยขอบแยกกัน แสดงดังตาราง 1, 2 และ 3

ตาราง 1 แสดงค่าความคลาดเคลื่อนกำลังสองเฉลี่ย จำแนกตามระดับความสัมพันธ์ และขนาดพื้นที่ย่อย ภายใต้ประชากรสถานการณ์ที่ 1 กรณีหน่วยขอบแยกกัน 2 กลุ่ม

ระดับความสัมพันธ์	n	วิธีประมาณค่าเฉลี่ย		
		SE	RA	RB
ต่ำ	3	1.0112	<u>0.6755</u>	0.6779
	4	<u>0.7229</u>	0.7331	0.7403
	5	<u>0.6649</u>	0.7088	0.7142
	6	<u>0.5791</u>	0.7079	0.7311
	7	<u>0.4821</u>	0.6945	0.7130
	8	<u>0.3339</u>	0.6871	0.7221
	9	<u>0.3966</u>	0.6687	0.7097
	10	<u>0.3657</u>	0.6292	0.6910
	11	<u>0.3341</u>	0.6211	0.6692
	12	<u>0.3017</u>	0.5442	0.6179
	16	<u>0.2600</u>	0.4761	0.5679
	20	<u>0.2348</u>	0.3920	0.5273
	24	<u>0.2236</u>	0.3195	0.4999
ปานกลาง	3	1.1777	0.2743	<u>0.2710</u>
	4	0.6425	0.2436	<u>0.2324</u>
	5	0.5549	0.2345	<u>0.2309</u>
	6	0.5281	0.2199	<u>0.2123</u>
	7	0.5002	0.2012	<u>0.1951</u>
	8	0.4394	0.1889	<u>0.1810</u>

ตาราง 1 (ต่อ)

ระดับความสัมพันธ์	*n*	วิธีประมาณค่าเฉลี่ย		
		SE	RA	RB
ปานกลาง	9	0.4085	0.1711	**0.1637**
	10	0.3959	0.1644	**0.1563**
	11	0.3648	0.1514	**0.1437**
	12	0.3296	0.1369	**0.1362**
	16	0.2549	0.1148	**0.1086**
	20	0.2318	**0.0911**	0.0929
	24	0.2126	**0.0793**	0.0813
สูง	3	1.1175	0.1914	**0.1910**
	4	0.8049	0.1668	**0.1662**
	5	0.7456	0.1424	**0.1413**
	6	0.6106	0.1279	**0.1267**
	7	0.5465	0.1002	**0.0987**
	8	0.4527	0.0926	**0.0907**
	9	0.4049	0.0891	**0.0852**
	10	0.3823	0.0712	**0.0679**
	11	0.3501	0.0621	**0.0600**
	12	0.3175	0.0566	**0.0541**
	16	0.2539	0.0264	**0.0229**
	20	0.2268	0.0185	**0.0158**
	24	0.2045	0.0123	**0.0103**

หมายเหตุ: ตัวหนา และขีดเส้นใต้ แทน ความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุดในแต่ละสถานการณ์

จากตาราง 1 แสดงความคลาดเคลื่อนกำลังสองเฉลี่ย ที่**ระดับความสัมพันธ์ต่ำ** เมื่อ $n=3$ พบว่า วิธี RA มีความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุด รองลงมาได้แก่วิธี RB และ SE ตามลำดับ แต่เมื่อ $n=4$ ขึ้นไป พบว่า วิธี SE มีความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุดรองลงมาได้แก่วิธี RA และ RB ตามลำดับ สำหรับ**ระดับความสัมพันธ์ปานกลาง** เมื่อ $n=3$ขึ้นไป พบว่า วิธี RB มีความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุด รองลงมาได้แก่วิธี RA และ SE ตามลำดับ ยกเว้น เมื่อ $n=20$ และ 24 พบว่า วิธี RA มีความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุด รองลงมาได้แก่วิธี RB และ SE ตามลำดับ และที่**ระดับความสัมพันธ์สูง** เมื่อ $n=3$ ขึ้นไป พบว่า วิธี RB มีความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุด รองลงมาได้แก่วิธี RA และ SE ตามลำดับ

นอกจากนี้เมื่อขนาดพื้นที่ย่อยเพิ่มขึ้น ความคลาดเคลื่อนกำลังสองเฉลี่ยของวิธีประมาณทั้งสามวิธีมีแนวโน้มลดลง ยกเว้นกรณีระดับความสัมพันธ์ต่ำ ขนาดตัวอย่างเท่ากับ 4, 6 และ 8 ของวิธีการประมาณ RA และ RB

ตาราง 2 แสดงค่าความคลาดเคลื่อนกำลังสองเฉลี่ย จำแนกตามระดับความสัมพันธ์ และขนาดพื้นที่ย่อย ภายใต้ประชากรสถานการณ์ที่ 1 กรณีหน่วยขอบแยกกัน 3 กลุ่ม

ระดับความสัมพันธ์	n	วิธีประมาณค่าเฉลี่ย		
		SE	RA	RB
ต่ำ	3	1.3625	**1.0627**	1.0775
	4	**1.0061**	1.0495	1.0737
	5	**0.8904**	1.0553	1.0649
	6	**0.6921**	1.0958	1.1320
	7	**0.6295**	1.0617	1.1251
	8	**0.5138**	1.0437	1.1119
	9	**0.4859**	0.9675	1.0504
	10	**0.4610**	0.9149	1.0450
	11	**0.4426**	0.8925	1.0603

ตาราง 2 (ต่อ)

ระดับความสัมพันธ์	n	วิธีประมาณค่าเฉลี่ย		
		SE	RA	RB
ต่ำ	12	0.4047	0.8566	1.0087
	16	0.3579	0.7479	0.9680
	20	0.2845	0.6103	0.9059
	24	0.2611	0.5014	0.8260
ปานกลาง	3	1.2892	0.4069	0.4106
	4	1.0390	0.4005	0.4072
	5	0.9245	0.3995	0.4088
	6	0.7135	0.3487	0.3608
	7	0.6567	0.3503	0.3720
	8	0.5585	0.3292	0.3469
	9	0.4713	0.3150	0.3381
	10	0.4978	0.2938	0.3168
	11	0.4095	0.2556	0.2775
	12	0.3798	0.2449	0.2728
	16	0.3560	0.1917	0.2269
	20	0.2897	0.1679	0.2110
	24	0.2715	0.1438	0.1878
สูง	3	1.3479	0.2211	0.2208
	4	1.0736	0.1759	0.1753
	5	0.8536	0.1560	0.1553
	6	0.6793	0.1340	0.1329
	7	0.6127	0.1133	0.1117
	8	0.5400	0.0914	0.0898

UTTARADIT RAJABHAT UNIVERSITY มหาวิทยาลัยราชภัฏอุตรดิตถ์

ตาราง 2 (ต่อ)

ระดับความสัมพันธ์	n	วิธีประมาณค่าเฉลี่ย		
		SE	RA	RB
สูง	9	0.5139	0.0726	**<u>0.0708</u>**
	10	0.4655	0.0678	**<u>0.0655</u>**
	11	0.4407	0.0564	**<u>0.0542</u>**
	12	0.4199	0.0509	**<u>0.0487</u>**
	16	0.3580	0.0249	**<u>0.0222</u>**
	20	0.2960	0.0158	**<u>0.0132</u>**
	24	0.2735	0.0094	**<u>0.0070</u>**

หมายเหตุ: ตัวหนา และขีดเส้นใต้ แทน ความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุดในแต่ละสถานการณ์

จากตาราง 2 แสดงความคลาดเคลื่อนกำลังสองเฉลี่ย ที่**ระดับความสัมพันธ์ต่ำ** เมื่อ $n=3$ พบว่าวิธี RA มีความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุด รองลงมาได้แก่วิธี RB และ SE ตามลำดับ แต่เมื่อ $n=4$ ขึ้นไป พบว่า วิธี SE มีความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุดรองลงมาได้แก่วิธี RA และ RB ตามลำดับ สำหรับ**ระดับความสัมพันธ์ปานกลาง** เมื่อ $n=3$ ขึ้นไป พบว่า วิธี RA มีความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุด รองลงมาได้แก่วิธี RB และ SE ตามลำดับ และที่**ระดับความสัมพันธ์สูง** เมื่อ $n=3$ ขึ้นไป พบว่า วิธี RB มีความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุด รองลงมาได้แก่วิธี RA และ SE ตามลำดับ

นอกจากนี้เมื่อขนาดพื้นที่ย่อยเพิ่มขึ้น ความคลาดเคลื่อนกำลังสองเฉลี่ยของวิธีประมาณทั้งสามวิธีมีแนวโน้มลดลง ยกเว้นกรณีระดับความสัมพันธ์ต่ำ ขนาดพื้นที่ย่อยเท่ากับ 5 และ 6 ของวิธีการประมาณ RA ขนาดพื้นที่ย่อยเท่ากับ 6 และ 11 ของวิธีการ RB สำหรับระดับความสัมพันธ์ปานกลาง ขนาดพื้นที่ย่อยเท่ากับ 7 ของวิธีการประมาณ RA ขนาดพื้นที่ย่อยเท่ากับ 5 และ 7 ของวิธีการประมาณ RB ขนาดพื้นที่ย่อยเท่ากับ 10 ของวิธีการประมาณ SE

ตาราง 3 แสดงค่าความคลาดเคลื่อนกำลังสองเฉลี่ย จำแนกตามระดับความสัมพันธ์ และขนาดพื้นที่ย่อย ภายใต้ประชากรสถานการณ์ที่ 1 กรณีหน่วยขอบแยกกัน 4 กลุ่ม

ระดับความสัมพันธ์	*n*	วิธีประมาณค่าเฉลี่ย		
		SE	RA	RB
ต่ำ	3	**2.3274**	2.5300	2.4948
	4	**2.1516**	2.3614	2.3268
	5	**1.7505**	2.1683	2.1027
	6	**1.5326**	2.0838	2.0069
	7	**1.2444**	1.6451	1.6067
	8	**1.1582**	1.3578	1.2987
	9	**1.0815**	1.2905	1.2327
	10	**1.0088**	1.2523	1.2035
	11	**0.9381**	0.9722	0.9671
	12	**0.8293**	0.9736	0.9715
	16	0.7928	**0.6827**	0.7523
	20	0.7727	**0.5041**	0.6447
	24	0.7068	**0.4222**	0.5939
ปานกลาง	3	2.9206	1.6676	**1.6643**
	4	2.2815	**1.7788**	1.8079
	5	1.8461	1.3520	**1.3303**
	6	1.5215	**1.2640**	1.3073
	7	1.3685	**0.8481**	0.8585
	8	1.2190	0.6322	**0.6321**
	9	1.0422	0.7120	**0.7069**

ตาราง 3 (ต่อ)

ระดับความสัมพันธ์	*n*	วิธีประมาณค่าเฉลี่ย		
		SE	RA	RB
ปานกลาง	10	1.0195	0.7426	**<u>0.7392</u>**
	11	0.9015	0.6329	**<u>0.6119</u>**
	12	0.8600	0.4355	**<u>0.4065</u>**
	16	0.8039	0.3095	**<u>0.3066</u>**
	20	0.7698	0.1806	**<u>0.1629</u>**
	24	0.8099	0.1320	**<u>0.1260</u>**
สูง	3	2.8950	0.6098	**<u>0.6061</u>**
	4	2.3074	0.4175	**<u>0.4119</u>**
	5	1.7689	0.3306	**<u>0.3252</u>**
	6	1.5649	0.2513	**<u>0.2457</u>**
	7	1.3257	0.2302	**<u>0.2228</u>**
	8	1.2238	0.1341	**<u>0.1248</u>**
	9	1.0721	0.1083	**<u>0.1016</u>**
	10	0.9865	0.0748	**<u>0.0667</u>**
	11	0.9710	0.0562	**<u>0.0482</u>**
	12	0.8664	0.0498	**<u>0.0391</u>**
	16	0.8058	0.0217	**<u>0.0157</u>**
	20	0.7763	0.0101	**<u>0.0061</u>**
	24	0.8132	0.0056	**<u>0.0024</u>**

หมายเหตุ: ตัวหนา และขีดเส้นใต้ แทน ความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุดในแต่ละสถานการณ์

จากตาราง 3 แสดงความคลาดเคลื่อนกำลังสองเฉลี่ย ที่**ระดับความสัมพันธ์ต่ำ** เมื่อ เมื่อ $n=3$ ขึ้นไป พบว่า วิธี SE มีความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุดรองลงมาได้แก่วิธี RB และ RA ตามลำดับ ยกเว้น เมื่อ $n=16,\ 20$ และ 24 พบว่าวิธี RA มีความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุด

รองลงมาได้แก่วิธี RB และ SE ตามลำดับ สำหรับ**ระดับความสัมพันธ์ปานกลาง** โดยทั่วไปวิธี RB มีความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุด รองลงมาได้แก่วิธี RA และ SE ตามลำดับ ยกเว้นเมื่อ $n = 4, 6$ และ 7 วิธี RA มีความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุด รองลงมาได้แก่วิธี RB และ SE ตามลำดับ และเมื่อขนาด $n = 3$ และ 5 วิธี RB มีความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุด รองลงมาได้แก่วิธี RA และ SE ตามลำดับ และที่**ระดับความสัมพันธ์สูง** เมื่อ $n = 3$ ขึ้นไป พบว่า วิธี RB มีความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุด รองลงมาได้แก่วิธี RA และ SE ตามลำดับ

นอกจากนี้เมื่อขนาดพื้นที่ย่อยเพิ่มขึ้น ความคลาดเคลื่อนกำลังสองเฉลี่ยของวิธีประมาณทั้งสามวิธีมีแนวโน้มลดลง ยกเว้นกรณีระดับความสัมพันธ์ต่ำ ขนาดพื้นที่ย่อยเท่ากับ 12 ของวิธีการ RA และ RB สำหรับระดับความสัมพันธ์ปานกลาง ขนาดพื้นที่ย่อยเท่ากับ 4 และ 9 ของวิธีการประมาณ RA และ RB ขนาดพื้นที่ย่อยเท่ากับ 24 ของวิธีการประมาณ SE และที่ระดับความสัมพันธ์สูงขนาดพื้นที่ย่อยเท่ากับ 24 ของวิธีการประมาณ SE

ตอนที่ 2 ผลการเปรียบเทียบความคลาดเคลื่อนกำลังสองเฉลี่ย ภายใต้ประชากรสถานการณ์ที่ 2 กรณีหน่วยขอบร่วมกัน แสดงดังตาราง 4, 5 และ 6

ตาราง 4 แสดงค่าความคลาดเคลื่อนกำลังสองเฉลี่ย จำแนกตามระดับความสัมพันธ์ และขนาดพื้นที่ย่อย ภายใต้ประชากรสถานการณ์ที่ 2 กรณีหน่วยขอบร่วมกัน 2 กลุ่ม

ระดับความสัมพันธ์	n	วิธีประมาณค่าเฉลี่ย		
		SE	RA	RB
ต่ำ	3	1.5563	1.0698	**<u>1.0612</u>**
	4	1.2628	1.0600	**<u>1.0528</u>**
	5	**<u>0.9882</u>**	1.0473	1.0300

ตาราง 4 (ต่อ)

ระดับความสัมพันธ์	*n*	วิธีประมาณค่าเฉลี่ย		
		SE	RA	RB
ต่ำ	6	**0.9196**	0.9612	0.9570
	7	**0.7327**	0.8802	0.8778
	8	**0.6961**	0.9084	0.8911
	9	**0.5943**	0.8228	0.8213
	10	**0.5807**	0.7306	0.7364
	11	**0.5502**	0.6815	0.7171
	12	**0.5057**	0.6479	0.6700
	16	**0.4579**	0.4975	0.5944
	20	0.4297	**0.3984**	0.5308
	24	0.4524	**0.3274**	0.5093
ปานกลาง	3	1.6053	0.3766	**0.3719**
	4	1.1925	0.3213	**0.3137**
	5	0.9919	0.2922	**0.2821**
	6	0.7823	0.2489	**0.2367**
	7	0.7472	0.2150	**0.1998**
	8	0.7011	0.1913	**0.1724**
	9	0.6156	0.1779	**0.1622**
	10	0.5634	0.1511	**0.1327**
	11	0.5010	0.1264	**0.1019**
	12	0.5254	0.1263	**0.1031**
	16	0.4512	0.0797	**0.0623**
	20	0.4209	0.0624	**0.0486**
	24	0.4461	0.0505	**0.0424**

ตาราง 4 (ต่อ)

ระดับความสัมพันธ์	n	วิธีประมาณค่าเฉลี่ย		
		SE	RA	RB
สูง	3	1.5542	0.3212	**<u>0.3205</u>**
	4	1.2550	0.2588	**<u>0.2572</u>**
	5	0.9957	0.2126	**<u>0.2108</u>**
	6	0.8516	0.1821	**<u>0.1794</u>**
	7	0.7523	0.1455	**<u>0.1427</u>**
	8	0.6604	0.1154	**<u>0.1122</u>**
	9	0.6321	0.1061	**<u>0.1032</u>**
	10	0.5424	0.0796	**<u>0.0758</u>**
	11	0.5579	0.0703	**<u>0.0670</u>**
	12	0.4853	0.0556	**<u>0.0513</u>**
	16	0.4579	0.0261	**<u>0.0225</u>**
	20	0.4309	0.0159	**<u>0.0129</u>**
	24	0.4560	0.0091	**<u>0.0071</u>**

หมายเหตุ: ตัวหนา และขีดเส้นใต้ แทน ความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุดในแต่ละสถานการณ์

จากตาราง 4 แสดงความคลาดเคลื่อนกำลังสองเฉลี่ย ที่**ระดับความสัมพันธ์ต่ำ** เมื่อ $n=3$ และ 4 พบว่า วิธี RB มีความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุด รองลงมาได้แก่วิธี RA และ SE ตามลำดับ แต่เมื่อ $n=$ 5 – 9 พบว่า วิธี SE มีความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุด รองลงมาได้แก่วิธี RB และ RA ตามลำดับ และเมื่อ $n=10-16$ วิธี SE มีความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุด ได้แก่วิธี RA และ RB ตามลำดับ และเมื่อ $n=20$ และ 24 วิธี RA มีความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุด รองลงมาได้แก่วิธี SE และ RB ตามลำดับ สำหรับ**ระดับความสัมพันธ์ปานกลาง** เมื่อ $n=3$ ขึ้นไป พบว่า วิธี RB มีความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุด รองลงมาได้แก่วิธี RA และ SE ตามลำดับ และที่**ระดับความสัมพันธ์สูง** เมื่อ $n=3$ ขึ้นไป พบว่า วิธี RB มีความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุดรองลงมาได้แก่วิธี RA และ SE ตามลำดับ

นอกจากนี้ขนาดพื้นที่ย่อยเพิ่มขึ้น ความคลาดเคลื่อนกำลังสองเฉลี่ยของวิธีประมาณทั้งสามวิธีมีแนวโน้มลดลง ยกเว้นกรณีระดับความสัมพันธ์ต่ำ ขนาดพื้นที่ย่อยเท่ากับ 8 ของวิธีการ RA และ RB ขนาดพื้นที่ย่อยเท่ากับ 24 ของวิธีการประมาณ SE สำหรับระดับความสัมพันธ์ปานกลาง ขนาดพื้นที่ย่อยเท่ากับ 12 และ 24 ของวิธีการประมาณ SE และที่ระดับความสัมพันธ์สูงขนาดพื้นที่ย่อยเท่ากับ 11 และ 24 ของวิธีการประมาณ SE

ตาราง 5 แสดงค่าความคลาดเคลื่อนกำลังสองเฉลี่ย จำแนกตามระดับความสัมพันธ์ และขนาดพื้นที่ย่อย ภายใต้ประชากรสถานการณ์ที่ 2 กรณีหน่วยขอบร่วมกัน 3 กลุ่ม

ระดับความสัมพันธ์	n	วิธีประมาณค่าเฉลี่ย		
		SE	RA	RB
ต่ำ	3	**2.7870**	2.8484	2.9046
	4	**2.6685**	2.8607	2.9017
	5	**1.9965**	2.7579	2.8678
	6	**1.7513**	2.4567	2.6355
	7	**1.5581**	2.3937	2.5752
	8	**1.4358**	1.8362	2.0680
	9	**1.2489**	1.8033	2.0390
	10	**1.1624**	1.7026	1.9859
	11	**1.0945**	1.5035	1.8607
	12	**1.0448**	1.2789	1.5233
	16	**0.8937**	0.8948	1.2039
	20	0.8311	**0.6821**	1.0262
	24	0.8526	**0.5859**	0.9610

ตาราง 5 (ต่อ)

ระดับความสัมพันธ์	*n*	วิธีประมาณค่าเฉลี่ย		
		SE	RA	RB
ปานกลาง	3	3.3600	1.0937	<u>**1.0864**</u>
	4	2.4496	0.9740	<u>**0.9545**</u>
	5	2.2052	0.8279	<u>**0.8156**</u>
	6	1.7037	0.7382	<u>**0.7211**</u>
	7	1.4750	0.6755	<u>**0.6514**</u>
	8	1.2825	0.6214	<u>**0.6021**</u>
	9	1.3439	0.5671	<u>**0.5483**</u>
	10	1.0566	0.5005	<u>**0.4693**</u>
	11	1.1164	0.4933	<u>**0.4928**</u>
	12	0.9739	<u>**0.4517**</u>	0.4523
	16	0.9256	<u>**0.3543**</u>	0.3788
	20	0.8232	<u>**0.2821**</u>	0.3278
	24	0.8543	<u>**0.2419**</u>	0.3270
สูง	3	3.4453	0.5238	<u>**0.5226**</u>
	4	2.5309	0.3831	<u>**0.3816**</u>
	5	2.0723	0.2689	<u>**0.2663**</u>
	6	1.8450	0.1830	<u>**0.1796**</u>
	7	1.4311	0.1614	<u>**0.1578**</u>
	8	1.2653	0.1108	<u>**0.1066**</u>
	9	1.2028	0.0786	<u>**0.0741**</u>
	10	1.1945	0.0506	<u>**0.0459**</u>
	11	1.0072	0.0389	<u>**0.0343**</u>
	12	1.0202	0.0394	<u>**0.0350**</u>

ตาราง 5 (ต่อ)

ระดับความสัมพันธ์	n	วิธีประมาณค่าเฉลี่ย		
		SE	RA	RB
สูง	16	0.9289	0.0114	**0.0084**
	20	0.8332	0.0085	**0.0053**
	24	0.8573	0.0048	**0.0026**

หมายเหตุ: ตัวหนา และขีดเส้นใต้ แทน ความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุดในแต่ละสถานการณ์

จากตาราง 5 แสดงความคลาดเคลื่อนกำลังสองเฉลี่ย **ที่ระดับความสัมพันธ์ต่ำ** เมื่อ เมื่อ $n=3$ ขึ้นไป พบว่า วิธี SE มีความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุดรองลงมาได้แก่วิธี RA และ RB ตามลำดับ ยกเว้น เมื่อ $n=20$ และ 24 พบว่าวิธี RA มีความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุด รองลงมาได้แก่วิธี SE และ RB ตามลำดับ สำหรับ**ระดับความสัมพันธ์ปานกลาง** เมื่อ $n=3-11$ พบว่า วิธี RB มีความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุดรองลงมาได้แก่วิธี RA และ SE ตามลำดับ และเมื่อ $n=12$ ขึ้นไป พบว่าวิธี RA มีความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุด รองลงมาได้แก่วิธี RB และ SE ตามลำดับ และ**ที่ระดับความสัมพันธ์สูง** เมื่อ $n=3$ ขึ้นไป พบว่า วิธี RB มีความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุดรองลงมาได้แก่วิธี RA และ SE ตามลำดับ

นอกจากนี้เมื่อขนาดพื้นที่ย่อยเพิ่มขึ้น ความคลาดเคลื่อนกำลังสองเฉลี่ยของวิธีประมาณทั้งสามวิธีมีแนวโน้มลดลง ยกเว้นกรณีระดับความสัมพันธ์ต่ำ ขนาดพื้นที่ย่อยเท่ากับ 4 ของวิธีการ RA ขนาดพื้นที่ย่อยเท่ากับ 24 ของวิธีการประมาณ SE สำหรับระดับความสัมพันธ์ปานกลาง ขนาดพื้นที่ย่อยเท่ากับ 9, 11 และ 24 ของวิธีการประมาณ SE ขนาดพื้นที่ย่อยเท่ากับ 11 ของวิธีการประมาณ RB และที่ระดับความสัมพันธ์สูงขนาดพื้นที่ย่อยเท่ากับ 12 ของวิธีการประมาณ SE, RA และ RB ขนาดพื้นที่ย่อยเท่ากับ 24 ของวิธีการประมาณ SE

ตาราง 6 แสดงค่าความคลาดเคลื่อนกำลังสองเฉลี่ย จำแนกตามระดับความสัมพันธ์ และขนาดพื้นที่ย่อย ภายใต้ประชากรสถานการณ์ที่ 2 กรณีหน่วยขอบร่วมกัน 4 กลุ่ม

ระดับความสัมพันธ์	n	วิธีประมาณค่าเฉลี่ย		
		SE	RA	RB
ต่ำ	3	1.9040	1.5603	**1.5214**
	4	1.6963	**1.4644**	1.4824
	5	1.4097	1.3695	**1.3438**
	6	**1.1454**	1.3065	1.2876
	7	**1.0133**	1.1369	1.1044
	8	**0.8606**	1.0194	0.9712
	9	**0.8139**	0.9620	0.9416
	10	**0.7597**	0.9138	0.9114
	11	**0.6833**	0.7950	0.8099
	12	**0.6558**	0.7626	0.7967
	16	**0.5467**	0.5571	0.6813
	20	0.5111	**0.4339**	0.5768
	24	0.5267	**0.3678**	0.5559
ปานกลาง	3	2.2237	0.6750	**0.6577**
	4	1.6565	0.5654	**0.5396**
	5	1.3409	0.4895	**0.4585**
	6	1.1398	0.4060	**0.3644**
	7	0.9526	0.3868	**0.3418**
	8	0.8503	0.3596	**0.3176**
	9	0.7742	0.2856	**0.2426**

ตาราง 6 (ต่อ)

ระดับความสัมพันธ์	n	วิธีประมาณค่าเฉลี่ย		
		SE	RA	RB
ปานกลาง	10	0.7102	0.2792	**<u>0.2167</u>**
	11	0.7042	0.2572	**<u>0.2056</u>**
	12	0.6503	0.2306	**<u>0.1818</u>**
	16	0.5433	0.1678	**<u>0.1290</u>**
	20	0.5141	0.1257	**<u>0.0960</u>**
	24	0.5078	0.1064	**<u>0.0928</u>**
สูง	3	2.1716	0.4286	**<u>0.4268</u>**
	4	1.6302	0.3395	**<u>0.3362</u>**
	5	1.4014	0.2504	**<u>0.2273</u>**
	6	1.1520	0.1821	**<u>0.1786</u>**
	7	0.9427	0.1680	**<u>0.1510</u>**
	8	0.8059	0.1071	**<u>0.1026</u>**
	9	0.7284	0.0853	**<u>0.0719</u>**
	10	0.7080	0.0753	**<u>0.0678</u>**
	11	0.6650	0.0559	**<u>0.0506</u>**
	12	0.6486	0.0482	**<u>0.0425</u>**
	16	0.5616	0.0164	**<u>0.0095</u>**
	20	0.5121	0.0073	**<u>0.0027</u>**
	24	0.4999	0.0047	**<u>0.0011</u>**

หมายเหตุ: ตัวหนา และขีดเส้นใต้ แทน ความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุดในแต่ละสถานการณ์

จากตาราง 6 แสดงความคลาดเคลื่อนกำลังสองเฉลี่ย ที่**ระดับความสัมพันธ์ต่ำ** เมื่อ $n=3$ และ 5 พบว่า วิธี RB มีความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุด รองลงมาได้แก่วิธี RA และ SE ตามลำดับ $n=4$ พบว่า วิธี RA มีความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุด รองลงมาได้แก่วิธี RB และ

SE ตามลำดับ $n=6-10$ พบว่า วิธี SE มีความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุด รองลงมาได้แก่วิธี RB และ RA ตามลำดับ $n=11-16$ พบว่า วิธี SE มีความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุด รองลงมาได้แก่วิธี ได้แก่วิธี RA และ RB ตามลำดับ และเมื่อ $n=20$ และ 24 พบว่าวิธี RA มีความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุดรองลงมาได้แก่วิธี SE และ RB ตามลำดับ สำหรับ**ระดับความสัมพันธ์ปานกลาง** เมื่อ $n=3$ ขึ้นไป พบว่า วิธี RB มีความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุด รองลงมาได้แก่วิธี RA และ SE ตามลำดับ และที่**ระดับความสัมพันธ์สูง** เมื่อ $n=3$ ขึ้นไป พบว่า วิธี RB มีความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุด รองลงมาได้แก่วิธี RA และ SE ตามลำดับ

นอกจากนี้เมื่อขนาดพื้นที่ย่อยเพิ่มขึ้น ความคลาดเคลื่อนกำลังสองเฉลี่ยของวิธีประมาณทั้งสามวิธีมีแนวโน้มลดลง ยกเว้นกรณีระดับความสัมพันธ์ต่ำ ขนาดพื้นที่ย่อยเท่ากับ 24 ของวิธีการประมาณ SE

ตอนที่ 3 สรุปจำนวนความถี่วิธีประมาณค่าเฉลี่ย จำแนกตามระดับความสัมพันธ์ **จำนวนหน่วยขอบ และ**สถานการณ์ที่หน่วยขอบแยกกัน และหน่วยขอบร่วมกัน

ตาราง 7 แสดงสรุปจำนวนความถี่วิธีประมาณค่าเฉลี่ย จำแนกตามระดับความสัมพันธ์ จำนวนหน่วยขอบ และสถานการณ์ที่หน่วยขอบแยกกัน และหน่วยขอบร่วมกัน

ระดับความสัมพันธ์	สถานการณ์	จำนวนกลุ่ม	วิธีการประมาณค่า		
			SE	RA	RB
ต่ำ	หน่วยขอบแยก	2	12	1	0
		3	12	1	0
		4	10	3	0
	หน่วยขอบร่วม	2	9	2	2
		3	11	2	0
		4	8	3	2

ตาราง 7 (ต่อ)

ระดับความสัมพันธ์	สถานการณ์	จำนวนกลุ่ม	วิธีการประมาณค่า		
			SE	RA	RB
ปานกลาง	หน่วยขอบแยก	**2**	0	2	**<u>11</u>**
		3	0	**<u>13</u>**	0
		4	0	3	**<u>10</u>**
	หน่วยขอบร่วม	**2**	0	0	**<u>13</u>**
		3	0	4	**<u>9</u>**
		4	0	0	**<u>13</u>**
สูง	หน่วยขอบแยก	**2**	0	0	**<u>13</u>**
		3	0	0	**<u>13</u>**
		4	0	0	**<u>13</u>**
	หน่วยขอบร่วม	**2**	0	0	**<u>13</u>**
		3	0	0	**<u>13</u>**
		4	0	0	**<u>13</u>**

หมายเหตุ: ตัวหนา และขีดเส้นใต้ แทน จำนวนความถี่ที่สุดในแต่ละสถานการณ์

จากตาราง 7 แสดงจำนวนความถี่วิธีประมาณค่าเฉลี่ย ที่**ระดับความสัมพันธ์ต่ำ**กรณีที่หน่วยขอบแยกกัน และหน่วยขอบร่วมกัน 2, 3 และ 4 กลุ่ม พบว่า วิธีการประมาณค่าอย่างง่ายมีความถี่สูงที่สุด สำหรับที่**ระดับความสัมพันธ์ปานกลาง** กรณีที่หน่วยขอบแยกกัน 2 และ 4 กลุ่ม และหน่วยขอบร่วมกัน 2, 3 และ 4 กลุ่ม พบว่า วิธีการประมาณค่าแบบอัตราส่วนที่ปรับปรุงโดยวิธีการ Rao-Blackwell มีความถี่สูงที่สุด ยกเว้นกรณีที่หน่วยขอบแยกกัน 3 กลุ่ม พบว่าวิธีการประมาณค่าแบบอัตราส่วนมีความถี่สูงที่สุด และที่**ระดับความสัมพันธ์สูง** กรณีที่หน่วยขอบแยกกันและหน่วยขอบร่วมกัน 2, 3 และ 4 กลุ่ม พบว่า วิธีการประมาณค่าแบบอัตราส่วนที่ปรับปรุงโดยวิธีการ Rao-Blackwell มีความถี่สูงที่สุด

บทที่ 5
บทสรุป

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพวิธีประมาณค่าเฉลี่ยที่ดัดแปลงมาจากตัวประมาณ Horvitz-Thompson ภายใต้แผนการสุ่มตัวอย่างแบบกลุ่มปรับด้วย วิธีประมาณค่าอย่างง่าย วิธีประมาณค่าแบบอัตราส่วน และวิธีประมาณค่าแบบอัตราส่วนที่ปรับปรุงโดยวิธีการ Rao-Blackwell เมื่อสุ่มหน่วยตัวอย่างแบบแบ่งชั้นภูมิ กรณีหน่วยขอบแยกกัน และกรณีหน่วยขอบร่วมกัน กำหนดขนาดพื้นที่ย่อยมีค่าเท่ากับ 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 16, 20 และ 24 หน่วย แบ่งความสัมพันธ์ระหว่างตัวแปรที่สนใจศึกษา และตัวแปรช่วย 3 ระดับ คือ ระดับต่ำ ปานกลาง และสูง เท่ากับ (0 , 0.4), [0.4, 0.7) และ [0.7, 1.00) ตามลำดับ ทำการจำลองข้อมูลในแต่ละสถานการณ์ทำซ้ำ 1,000 ครั้ง โดยใช้ความคลาดเคลื่อนกำลังสองเฉลี่ยเป็นเกณฑ์ในการเปรียบเทียบ ซึ่งวิธีประมาณค่าที่มีค่าความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุดจะเป็นวิธีประมาณค่าที่มีประสิทธิภาพมากที่สุด

สรุปผลการวิจัย

ผลการวิจัยสามารถสรุปได้โดยแยกออกเป็น 2 กรณี ดังนี้

1. กรณีหน่วยขอบแยกกัน

1.1 หน่วยขอบแยกกัน 2 กลุ่ม เมื่อพิจารณาทุกขนาดพื้นที่ย่อย และทุกระดับความสัมพันธ์ กรณีความสัมพันธ์อยู่ในระดับต่ำ พบว่า วิธีการประมาณค่าอย่างง่าย มีความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุดเป็นส่วนใหญ่ สำหรับกรณีความสัมพันธ์อยู่ในระดับปานกลาง และสูง พบว่า วิธีการประมาณค่าแบบอัตราส่วนที่ปรับปรุงโดยวิธีการ Rao-Blackwell มีความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุดเป็นส่วนใหญ่

1.2 หน่วยขอบแยกกัน 3 กลุ่ม เมื่อพิจารณาทุกขนาดพื้นที่ย่อย และทุกระดับความสัมพันธ์ กรณีความสัมพันธ์อยู่ในระดับต่ำ พบว่า วิธีการประมาณค่าอย่างง่าย มีความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุดเป็นส่วนใหญ่ สำหรับกรณีความสัมพันธ์อยู่ในระดับปานกลาง พบว่าวิธีการประมาณค่าแบบอัตราส่วนมีความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุดเป็นส่วนใหญ่ และกรณีความสัมพันธ์อยู่ในระดับสูง พบว่า วิธีการประมาณค่าแบบอัตราส่วนที่ปรับปรุงโดยวิธีการ Rao-Blackwell มีความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุดเป็นส่วนใหญ่

1.3 หน่วยขอบแยกกัน 4 กลุ่ม เมื่อพิจารณาทุกขนาดพื้นที่ย่อย และทุกระดับความสัมพันธ์ กรณีความสัมพันธ์อยู่ในระดับต่ำ พบว่า วิธีการประมาณค่าอย่างง่าย มีความคลาดเคลื่อน

เคลื่อนกำลังสองเฉลี่ยต่ำที่สุดเป็นส่วนใหญ่ สำหรับกรณีความสัมพันธ์อยู่ในระดับปานกลาง และสูง พบว่า วิธีการประมาณค่าแบบอัตราส่วนที่ปรับปรุงโดยวิธีการ Rao-Blackwell มีความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุดเป็นส่วนใหญ่

2. กรณีหน่วยขอบร่วมกัน

2.1 หน่วยขอบร่วมกัน 2 กลุ่ม เมื่อพิจารณาทุกขนาดพื้นที่ย่อย และทุกระดับความสัมพันธ์ กรณีความสัมพันธ์อยู่ในระดับต่ำ พบว่า วิธีการประมาณค่าอย่างง่าย มีความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุดเป็นส่วนใหญ่ สำหรับกรณีความสัมพันธ์อยู่ในระดับปานกลาง และสูง พบว่า วิธีการประมาณค่าแบบอัตราส่วนที่ปรับปรุงโดยวิธีการ Rao-Blackwell มีความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุดเป็นส่วนใหญ่

2.2 หน่วยขอบร่วมกัน 3 กลุ่ม เมื่อพิจารณาทุกขนาดพื้นที่ย่อย และทุกระดับความสัมพันธ์ กรณีความสัมพันธ์อยู่ในระดับต่ำ พบว่า วิธีการประมาณค่าอย่างง่าย มีความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุดเป็นส่วนใหญ่ สำหรับกรณีความสัมพันธ์อยู่ในระดับปานกลาง และสูง พบว่า วิธีการประมาณค่าแบบอัตราส่วนที่ปรับปรุงโดยวิธีการ Rao-Blackwell มีความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุดเป็นส่วนใหญ่

2.3 หน่วยขอบร่วมกัน 4 กลุ่ม เมื่อพิจารณาทุกขนาดพื้นที่ย่อย และทุกระดับความสัมพันธ์ กรณีความสัมพันธ์อยู่ในระดับต่ำ พบว่า วิธีการประมาณค่าอย่างง่าย มีความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุดเป็นส่วนใหญ่ สำหรับกรณีความสัมพันธ์อยู่ในระดับปานกลาง และสูง พบว่า วิธีการประมาณค่าแบบอัตราส่วนที่ปรับปรุงโดยวิธีการ Rao-Blackwell มีความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุดเป็นส่วนใหญ่

อภิปรายผลการวิจัย

จากการเปรียบเทียบวิธีประมาณค่าเฉลี่ยของทั้ง 2 กรณี พบว่าเมื่อทุกขนาดพื้นที่ย่อย และความสัมพันธ์อยู่ในระดับต่ำ วิธีการประมาณค่าอย่างง่ายเป็นวิธีที่เหมาะสมสำหรับการประมาณค่า เนื่องจากสูตรในการคำนวณหาค่าประมาณพารามิเตอร์ไม่ขึ้นอยู่กับตัวแปรช่วย จึงทำให้วิธีการประมาณค่าอย่างง่ายเป็นวิธีที่เหมาะสม ส่วนวิธีการประมาณค่าแบบอัตราส่วน สูตรในการคำนวณหาค่าประมาณพารามิเตอร์จะขึ้นอยู่ตัวแปรช่วย โดยเมื่อตัวแปรช่วย และตัวแปรที่สนใจศึกษามีความสัมพันธ์อยู่ในระดับต่ำ พบว่าค่า $\hat{R}u_{x_k}$ จะมีค่าน้อยกว่า u_{y_k} ส่งผลให้ค่า u'_{y_k} สูงขึ้น จึงทำให้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยมีค่าสูงขึ้น ในขณะเดียวกันเมื่อความสัมพันธ์ระหว่างตัวแปรช่วย และตัวแปรที่สนใจศึกษาอยู่ในระดับปานกลาง และสูง พบว่าค่า $\hat{R}u_{x_k}$ จะมีค่าใกล้เคียงกับค่า u_{y_k} จึงส่งผลให้ u'_{y_k} มีค่าน้อย ดังนั้นค่าความคลาดเคลื่อนกำลังสองเฉลี่ยจึงมีค่าลดน้อยลง และนอกจากนี้ผลการวิจัยยังพบว่าวิธีประมาณค่าแบบอัตราส่วนที่ปรับปรุงโดยวิธีการ Rao–Blackwell จะให้ค่า

ความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำกว่าวิธีการประมาณค่าแบบอัตราส่วน เนื่องจากค่าความแปรปรวนของวิธีประมาณค่าแบบอัตราส่วนที่ปรับปรุงโดยวิธีการ Rao–Blackwell ได้จากค่าความแปรปรวนของวิธีประมาณค่าแบบอัตราส่วนลบด้วยผลต่างกำลังสองของค่าเฉลี่ย จึงทำให้ค่าประมาณที่ได้มีความแปรปรวนต่ำกว่าวิธีการประมาณค่าแบบอัตราส่วน ซึ่งสอดคล้องกับผลการวิจัยของ Chao, Lin และ Chiang (2008) ที่ทำการศึกษา และเปรียบเทียบประสิทธิภาพตัวประมาณค่าเฉลี่ยแบบอัตราส่วนที่ปรับปรุงโดยวิธีการ Rao-Blackwell กับตัวประมาณอัตราส่วนแบบพื้นฐานผลการวิจัยพบว่า ตัวประมาณค่าเฉลี่ยแบบอัตราส่วนที่ปรับปรุงโดยวิธีการ Rao-Blackwell จะมีประสิทธิภาพมากกว่าตัวประมาณค่าเฉลี่ยอัตราส่วนแบบพื้นฐาน

เมื่อพิจารณาผลการวิจัยทั้ง 2 กรณี พบว่ากรณีหน่วยขอบร่วมกันหรือหน่วยขอบแยกกัน จะให้ผลที่สอดคล้องในทิศทางเดียวกัน เนื่องจากหน่วยขอบไม่ส่งผลต่อวิธีการประมาณค่า ดังนั้นผลการวิจัยที่ระดับความสัมพันธ์ต่ำ วิธีการประมาณค่าอย่างง่ายจะเป็นวิธีที่เหมาะสมสำหรับการประมาณค่า และเมื่อความสัมพันธ์อยู่ในระดับปานกลาง และสูง วิธีการประมาณค่าแบบอัตราส่วนที่ปรับปรุงโดยวิธีการ Rao–Blackwell เป็นวิธีที่เหมาะสมสำหรับการประมาณค่า ยกเว้นเมื่อหน่วยขอบแยกกัน 3 กลุ่ม ที่ระดับความสัมพันธ์ปานกลาง พบว่าวิธีการประมาณค่าแบบอัตราส่วน เป็นวิธีที่เหมาะสมสำหรับการประมาณค่า เนื่องจากการวิจัยนี้ระบุความสัมพันธ์ระหว่างตัวแปรเป็นช่วง ซึ่งอาจมีโอกาสที่ความสัมพันธ์ในระดับปานกลางและสูงจะมีค่าใกล้เคียงกัน ดังนั้นจึงส่งผลให้วิธีการประมาณค่าแบบอัตราส่วนมีค่าความคลาดเคลื่อนกำลังสองเฉลี่ยใกล้เคียงกับวิธีประมาณค่าแบบอัตราส่วนที่ปรับปรุงโดยวิธีการ Rao–Blackwell

นอกจากนี้ยังพบว่าเมื่อขนาดพื้นที่ย่อย และระดับความสัมพันธ์เพิ่มขึ้นความคลาดเคลื่อนกำลังสองเฉลี่ยของวิธีประมาณทั้งสามวิธีมีแนวโน้มลดลง เนื่องจากขนาดตัวอย่างที่มีค่าใกล้เคียงกับประชากรจะทำให้ค่าความเอนเอียงจากตัวอย่างลดน้อยลง จึงส่งผลให้การประมาณค่าทั้งสามวิธีมีประสิทธิภาพสูงขึ้น

ดังนั้นจึงกล่าวได้ว่าโดยทั่วไปวิธีการประมาณค่าอย่างง่ายเป็นวิธีที่เหมาะสมสำหรับการประมาณค่าพามิเตอร์เมื่อความสัมพันธ์อยู่ในระดับต่ำ สำหรับวิธีวิธีการประมาณค่าแบบอัตราส่วนที่ปรับปรุงโดย Rao–Blackwell เป็นวิธีที่เหมาะสมสำหรับการประมาณค่าพามิเตอร์เมื่อความสัมพันธ์อยู่ในระดับปานกลาง และสูง ทั้ง 2 กรณี

ข้อเสนอแนะ

1. ด้านการนำไปใช้ประโยชน์

การพิจารณาความสัมพันธ์ระหว่างตัวแปรที่สนใจศึกษา และตัวแปรช่วยสามารถพิจารณาได้จากค่าสัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน และจากผลการวิจัยสามารถนำไปใช้ประโยชน์ได้ดังนี้

การประมาณค่าเฉลี่ยของทั้ง 2 กรณี เมื่อทุกขนาดพื้นที่ย่อย และความสัมพันธ์อยู่ในระดับต่ำควรเลือกใช้วิธีการประมาณค่าอย่างง่าย แต่เมื่อความสัมพันธ์อยู่ในระดับปานกลาง และสูงควรเลือกใช้วิธีการประมาณค่าแบบอัตราส่วนที่ปรับปรุงโดยวิธีการ Rao-Blackwell เนื่องจากมีความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุดเป็นส่วนใหญ่

2. ด้านการศึกษาวิจัย

2.1 เป็นแนวทางในการศึกษาการประมาณค่าเฉลี่ยของแผนการสุ่มตัวอย่างแบบกลุ่มปรับ เมื่อสุ่มหน่วยตัวอย่างในรูปแบบอื่น เช่น และแผนการสุ่มตัวอย่างแบบมีระบบ เป็นต้น

2.2 เป็นแนวทางในการศึกษาการประมาณค่าเฉลี่ยของแผนการสุ่มตัวอย่างแบบกลุ่มปรับ เมื่อใช้วิธีประมาณค่าแบบอื่น เช่น วิธีการประมาณค่าแบบถดถอย เป็นต้น

UTTARADIT RAJABHAT UNIVERSITY

บรรณานุกรม

บรรณานุกรม

ทวีศักดิ์ ศิริพรไพบูลย์. (2549). **เทคนิคการสุ่มตัวอย่าง**. พิษณุโลก: มหาวิทยาลัยนเรศวร.

ประชุม สุวัตถี. (2545). **ทฤษฎีการอนุมานเชิงสถิติ**. (พิมพ์ครั้งที่ 2). กรุงเทพ ฯ: สำนักพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

ประชุม สุวัตถี. (2552). **การสำรวจด้วยตัวอย่าง: การชักตัวอย่างและการวิเคราะห์**. กรุงเทพ ฯ: สำนักงานกิจการโรงพิมพ์ องค์การสงเคราะห์ทหารผ่านศึก ในพระบรมราชูปถัมภ์.

วิชาญ โชควิวัฒน. (2546). **กรอบแนวคิดของแผนการสุ่มตัวอย่างแบบกลุ่มปรับ.** วิทยานิพนธ์ สต.ม., จุฬาลงกรณ์มหาวิทยาลัย, กรุงเทพ ฯ.

ศิรประภา มโนมัธย์. (2539). **ประสิทธิภาพของแผนการสุ่มตัวอย่างแบบกลุ่มปรับ เมื่อตัวอย่างขั้นต้นใช้วิธีการสุ่มอย่างง่าย แบบมีชั้นภูมิ และแบบมีระบบ ภายใต้แบบจำลองที่กำหนด**. วิทยานิพนธ์ วท.ม., มหาวิทยาลัยศิลปากร, กรุงเทพ ฯ.

สำนักงานสถิติแห่งชาติ. (พฤษภาคม 2542). **สรุปผลการสำรวจเบื้องต้นโครงการสำรวจภาวะการทำงานของประชากร รอบที่ 2 เดือนพฤษภาคม พ.ศ.2542**. สืบค้นเมื่อ 22 มิถุนายน 2561, จาก http://service.nso.go.th/nso/news/res42.htm.

สำนักงานสถิติแห่งชาติ. (กุมภาพันธ์ 2554). **โครงการสำรวจความคิดเห็นของประชาชนเกี่ยวกับสถานการณ์การแพร่ระบาดยาเสพย์ติด**. สืบค้นเมื่อ 22 มิถุนายน 2561, จาก http://service.nso.go.th/nso/nsopublish/servopin/serv_opinion.html

สุชาดา กีระนันท์. (2542). **ทฤษฎีและวิธีการสำรวจตัวอย่าง** (ครั้งที่ 2). กรุงเทพฯ: สำนักพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

Chao, C.T. (2004). Ratio estimation on adaptive cluster sampling. **Journal of Chinese Statistical Association**, 42(3), 141-180.

Chao, C.T., Lin, F.M. and Chiang, T.C. (2008). Improved ratio estimators in adaptive cluster sampling. **Section on Survey Research Methods**, 31(5), 3210-3216.

Cliff, A.D. and Ord, J.K. (1981). **Spatial processes models and applications**. London: Pion Limited.

Dryver, A.L. and Chao, C.T. (2007). Ratio estimators in adaptive cluster sampling. **Environmetric**, 18(6), 607-620.

Dryver, A.L. and Thompson, S.K. (2005). Improved unbiased estimators in adaptive cluster sampling. **Journal of the Royal Statistical Society B**, 67(1), 157-166.

Thompson, S.K. (1990), Adaptive Cluster Sampling. **Journal of the American Statistical Association**, 85(412), 1050-1059.

Thompson, S.K. (2002). **Sampling**. (2nd ed.). New York: Wiley.

Thompson, S.K. and Seber, G.A.F. (1996). **Adaptive Sampling**. New York: Wiley.
William G. Cochran. (1977). **Sampling Techniques.** (3rd ed.). New York: Wiley.

มหาวิทยาลัยราชภัฏอุตรดิตถ์
UTTARADIT RAJABHAT UNIVERSITY

ประวัติผู้วิจัย

หัวหน้าโครงการวิจัย

ชื่อ – นามสกุล นางสาววิไลวรรณ รัตนกูล
Miss.Wilaiwan Rattanakool

ตำแหน่งปัจจุบัน อาจารย์

หมายเลขบัตรประชาชน 1660390003622

หน่วยงาน หลักสูตรคณิตศาสตร์ คณะวิทยาศาสตร์และเทคโนโลยี
มหาวิทยาลัยราชภัฏอุตรดิตถ์
เลขที่ 27 ถ.อินใจมี ต.ท่าอิฐ อ.เมือง จ.อุตรดิตถ์ 53000
โทรศัพท์ 0-55416601 ต่อ 1300
โทรสาร 0-55416601 ต่อ 1312
E-mail : wilaiwan.ratt16@gmail.com

ประวัติการศึกษา วท.บ. สถิติ มหาวิทยาลัยนเรศวร ปี พ.ศ.ที่จบ 2550
วท.ม. สถิติประยุกต์ มหาวิทยาลัยนเรศวร ปี พ.ศ.ที่จบ 2554

ประสบการณ์วิจัย

ผู้ร่วมโครงการวิจัย : Hypothesis Testing for Difference of Mean of two Log-normal Distribution .Proceeding URU International Conference on Science & Technology 2016

งานวิจัยที่กำลังทำ

ผู้ร่วมโครงการวิจัย

ชื่อโครงการวิจัย : รูปแบบกิจกรรมการเรียนรู้คณิตศาสตร์สำหรับมัธยมศึกษาตอนต้น โรงเรียนขยายโอกาสตำบลป่าเซ่าและตำบลคุ้งตะเภา อ.เมือง จ.อุตรดิตถ์
Learning management model according mathematics for secondary education, Opportunity School in TumPasae and Kung Ta Pou District, Muang District, Uttaradit