บทที่ 1

บทนำ

1.1 ที่มาและความสำคัญของปัญหา

การสำรวจตัวอย่าง เป็นวิธีการเก็บรวบรวมข้อมูลตามสภาพที่เกิดขึ้นจริง โดยไม่มีการสร้างสถานการณ์ หรือแบบแผนเพื่อสร้างข้อมูลในลักษณะเฉพาะขึ้น ซึ่งจะเก็บข้อมูลจากหน่วยที่ให้ข้อมูลเพียงบางหน่วยเท่านั้น โดยวิธีการเลือกตัวอย่างที่เหมาะสม แทนที่จะจัดเก็บจากทุกหน่วยประชากร ทำให้สามารถลดทรัพยากร ทั้งเวลา กำลังคนและงบประมาณลงได้ และยังทำให้การวิเคราะห์ข้อมูลใช้เวลาน้อยลงหรือได้ผลการศึกษาเร็วขึ้น การรักษาและควบคุมคุณภาพในการจัดเก็บข้อมูลย่อมกระทำได้ง่ายขึ้น ทำให้ข้อมูลที่ได้มีความถูกต้องเชื่อถือได้มากขึ้น

หลักการพื้นฐานของการสำรวจตัวอย่าง คือ การเลือกตัวอย่างที่เป็นตัวแทนที่ดีของประชากร โดยคำนึงถึงลักษณะของประชากร และเงื่อนไข ข้อจำกัดต่างๆ เช่นกรอบสำหรับการเลือกตัวอย่าง หน่วยที่ให้ข้อมูลหรือหน่วยตัวอย่าง ทรัพยากรที่มี เป็นต้น แต่ข้อมูลที่ได้จะเป็นข้อมูลเพียงบางส่วน ไม่ใช่ข้อมูลที่สมบูรณ์ จึงต้องอาศัยวิธีการทางสถิติในการเลือกตัวอย่างจากประชากรและอนุมานกลับไปหาประชากรอีกชั้นหนึ่ง เพื่อสรุปผลหรือหาความรู้เกี่ยวกับประชากร รวมทั้งต้องพิจารณาระดับคุณภาพของข้อสรุปเกี่ยวกับประชากรที่ได้จากวิธีการเหล่านั้นด้วย

การเลือกตัวอย่างเชิงความน่าจะเป็น (probability sampling) เป็นแผนการเลือกตัวอย่างที่สามารถบอกความแม่นยำ (precision) ของตัวประมาณได้ นั่นคือจากประชากรขนาดจำกัดเท่ากับ N ประกอบด้วยหน่วยประชากร $y_1, y_2, ..., y_N$ ถ้าเลือกหน่วยตัวอย่างขนาด n $(y_1, y_2, ..., y_n)$ ด้วยความน่าจะเป็นที่หน่วย i จะถูกเลือกเข้ามาในตัวอย่างขนาด n (π_i) หน่วยในตัวอย่างอาจจะเป็นหน่วยใดก็ได้ในประชากร ไม่ใช่ว่าหน่วย i ในตัวอย่างคือหน่วย i ในประชากรค่า y_i เป็นค่าคงที่ที่ไม่ทราบของหน่วย i ในประชากร แต่ค่า y_i ในตัวอย่าง จะเป็นตัวแปรสุ่มเนื่องจาก ได้จากกระบวนการสุ่มซึ่งไม่สามารถคาดการณ์ได้ว่าผลการสุ่มตัวอย่างจะได้ตัวประมาณเป็นอย่างไรก่อนที่กระบวนการเลือกตัวอย่างจะสิ้นสุดลง แต่สามารถทราบได้ว่าผลของการสุ่มหรือชุดตัวอย่างที่เป็นไปได้มีอะไรบ้าง และเมื่อการเลือกตัวอย่างในลักษณะเดียวกันซ้ำๆกันหลายๆครั้ง แต่ละครั้งจะได้ชุดตัวอย่างขนาด n ที่เป็นไปได้ต่างๆกัน ด้วยความน่าจะเป็นของการเกิดชุดตัวอย่างเหล่านั้น ทำให้สามารถหาการแจกแจงความน่าจะเป็นของตัวแปรสุ่มหรือค่าต่างๆ ที่เป็นไปได้ของ

ตัวประมาณ และสามารถอาศัยวิธีการสถิติในการอธิบายคุณภาพของตัวประมาณได้ เช่น การเลือกตัวอย่างสุ่มแบบง่าย (simple random sampling ,srs) ซึ่งเป็นวิธีการเลือกตัวอย่างที่กำหนดให้ตัวอย่างในขนาดที่กำหนดทุกตัวอย่างที่อาจเป็นไปได้มีโอกาสเกิดขึ้นเท่าๆกัน เช่นจากประชากรขนาด N ถ้าต้องการเลือกตัวอย่างสุ่มแบบง่ายขนาด n เมื่อ N>n อย่างไม่ใส่คืน จะเห็นว่ามีตัวอย่างที่อาจเป็นไปได้ทั้งสิ้น $N!/\{(N-n)!n!\}$ ตัวอย่างแต่ละตัวอย่างที่เป็นไปได้นี้จะได้รับโอกาสถูกเลือกเท่าๆกันด้วยความน่าจะเป็น 1 / $N!/\{(N-n)!n!\}$ นั่นคือ ถ้า $\hat{y}_1, \hat{y}_2, ..., \hat{y}_{{}_NC_n}$ เป็นค่ายอดรวมของตัวอย่างขนาด n ที่เป็นไปได้ นำค่าเหล่านี้มาแจกแจงความถี่ จะได้ p_i เป็นความน่าจะเป็นในการเกิดค่า $\hat{y}_i$ นั่นคือ p_i เป็นการแจกแจงความน่าจะเป็นของ $\hat{y}_i$ (probability function)

ในบางครั้งค่าใช้จ่ายในการเดินทางระหว่าหน่วยตัวอย่างเพื่อทำการเก็บรวบรวมข้อมูล ผู้เลือกตัวอย่างก็อาจลดค่าใช้จ่ายการเดินทางลงโดยการเลือกตัวอย่างกลุ่ม (cluster sampling) เพื่อให้หน่วยตัวอย่างมีการกระจุกตัวแทนที่จะกระจายออกไปมาก ซึ่งค่าใช้จ่ายระหว่างหน่วยตัวอย่างภายในกลุ่มย่อมจะน้อยกว่ากรณีเลือกตัวอย่างโดยตรงจากกรอบตัวอย่าง หรือเพื่อแก้ปัญหาการขาดแคลนกรอบสำหรับเลือกตัวอย่าง (sampling frame) ซึ่งบางกรณีกรอบที่ประกอบด้วยรายชื่อหน่วยตัวอย่างที่ครบถ้วนและทันสมัยหาได้ยาก ทำให้การเลือกตัวอย่างแบบกลุ่มค่อนข้างแพร่หลายในการสำรวจตัวอย่างโดยทั่วไป การเลือกตัวอย่างแบบกลุ่ม เป็นการเลือกตัวอย่างที่หน่วยตัวอย่างมิใช่หน่วยที่ให้ข้อมูลโดยตรงเพียงระดับเดียว แต่เป็นหน่วยตัวอย่างที่ประกอบขึ้นด้วยหน่วยที่ให้ข้อมูลหลายๆหน่วยรวมกันเป็นกลุ่ม และกลุ่มอาจมีได้หลายระดับ โดยหน่วยตัวอย่างระดับถัดลงไปจะต้องเป็นหน่วยย่อย หรือหน่วยที่เล็กกว่าหน่วยตัวอย่างระดับก่อนหน้า (subunits or element) และจะเลือกเพียงบางกลุ่มเท่านั้น ซึ่งอาจเก็บข้อมูลจากทุกหน่วยในกลุ่มหรือเก็บเพียงบางหน่วยในกลุ่มก็ได้ ดังนั้นกลุ่มบางกลุ่มที่เลือกมาเป็นตัวอย่างจะต้องเป็นตัวแทนที่ดีของประชากรด้วย คือต้องเก็บลักษณะที่หลากหลายของประชากรไว้ในแต่ละกลุ่ม จึงจะทำให้ได้ตัวประมาณที่มีประสิทธิภาพ เช่น ประชากรขนาด N จะถูกแบ่งออกเป็นกลุ่ม ผู้เลือกตัวอย่างจะทำการเลือกกลุ่มตัวอย่างออกมาจำนวนหนึ่ง แล้วสร้างกรอบที่แสดงเฉพาะหน่วยต่างๆภายในกลุ่มตัวอย่างเหล่านั้น เพื่อทำการเลือกตัวอย่างระดับถัดไป ถ้าเป็นการเลือกกลุ่มตัวอย่างชั้นเดียวและให้แต่ละกลุ่มมีโอกาสถูกเลือกเท่าๆกัน หรือเลือกตัวอย่างสุ่มแบบง่ายขนาด n กลุ่มจาก N กลุ่มที่มี เมื่อได้ n กลุ่มแล้วจะเก็บรวบรวมข้อมูลจากทุกหน่วยที่อยู่ในกลุ่ม n กลุ่มนั้น และพิจารณาค่าประมาณเช่นค่ายอดรวมประชากรของกลุ่ม i เป็นเช่นเดียวกับค่าของหน่วย i ในกรณีของ srs

แผนการเลือกตัวอย่างเชิงความน่าจะเป็นที่เป็นสัดส่วนกับขนาด (probability proportional to size sampling , pps) เป็นการเลือกตัวอย่างเมื่อกลุ่มมีขนาดไม่เท่ากันหรือความสำคัญที่แต่ละกลุ่มมีไม่เท่าเทียมกัน เช่นถ้ากลุ่มคือจังหวัด จังหวัดบางจังหวัดมีจำนวนประชากรมากแตกต่างจากจังหวัดอื่น จึงควรได้รับโอกาสในการถูกเลือกสูงกว่า ซึ่งอาจมีผลทำให้เกิดความแม่นยำสูงขึ้นในการประมาณค่าประชากรได้ ในกรณีนี้ถือว่าจำนวนประชากรในจังหวัดจะเป็นค่าวัดขนาด (measure of size) ซึ่งค่าวัดขนาดไม่ได้หมายถึงขนาดของกลุ่มหรือจำนวนหน่วยภายในกลุ่มนั้นเพียงอย่างเดียว แต่อาจเป็นค่าวัดความสำคัญของจังหวัดหรือกลุ่มนั้น เช่นจำนวนเนื้อที่ถือครองทางการเกษตร ผลผลิตข้าวในปี 2540 ในระดับจังหวัด นั่นคือผู้เลือกตัวอย่างจะต้องเลือกเกณฑ์ที่มากำหนดโอกาสที่แต่ละกลุ่มจะถูกเลือก เช่น กำหนดด้วยขนาดของกลุ่ม หรือค่าแสดงความสำคัญของกลุ่มโดยเปรียบเทียบ ซึ่งอาจใช้ประโยชน์จาก ข้อมูลในอดีตหรือค่าของตัวแปรช่วยที่มีความสัมพันธ์กับตัวแปรที่สนใจศึกษามาใช้เป็นค่าวัดขนาด ส่วนการหาตัวประมาณค่ายอดรวมประชากร จะเป็นการปรับค่าในระดับตัวอย่างขึ้นไปสู่ระดับของประชากรโดยอาศัยความน่าจะเป็นที่ค่าตัวอย่างถูกเลือกมาจากประชากรนั้น ดังนั้นจึงต้องพิจารณาความน่าจะเป็นที่กลุ่ม i (π_i) ถูกเลือกเข้ามาในตัวอย่างขนาด n กลุ่ม (n>1) ที่เลือกจากประชากรขนาด N กลุ่ม แล้วทำการถ่วงน้ำหนักแต่ละหน่วยด้วยความน่าจะเป็น π_i นั่นคือตัวประมาณค่ายอดรวมประชากรคือ $\hat{y} = \sum_{i=1}^{n} \frac{y_i}{\pi_i}$ ซึ่งในกรณีของแผนการเลือกตัวอย่างแบบ pps ความน่าจะเป็น π_i จะเป็นสัดส่วนกับค่าของตัวแปร x นั่นคือ ความน่าจะเป็นที่กลุ่ม i ถูกเลือกเข้ามาในตัวอย่างเป็นกลุ่มหรือหน่วยแรกจากประชากรขนาด N กลุ่ม มีค่าเท่ากับ $p_i = \frac{x_i}{\sum_{i=1}^{N} x_i}$

ในกรณีที่ x เป็นตัวแปรปริมาณแต่อาจเป็นค่าที่ไม่น่าเชื่อถือ เช่น การเก็บข้อมูลเกี่ยวกับผลผลิตหรือรายได้ของเกษตรกรที่ให้ข้อมูลคลาดเคลื่อนเนื่องจากกลัวการเสียภาษีอากร หรืออาจเป็นการยากที่จะวัดค่าตัวแปร x จากทุกๆหน่วยในประชากรไม่คุ้มกับค่าใช้จ่ายหรือทรัพยากรที่ต้องเสียไป เช่น การวัดเส้นรอบวงของกิ่งไม้ เพื่อนำมากำหนดค่าความน่าจะเป็นในการเลือกกิ่งไม้ เพื่อประมาณจำนวนผลไม้บนต้นไม้ แต่ถ้าเราสามารถจัดอันดับของกิ่งไม้ได้ทุกๆกิ่ง หรือผลผลิตทางการเกษตรในจังหวัด การใช้ค่าอันดับของหน่วยประชากรเป็นค่าวัดขนาดมาประมาณค่ายอดรวมประชากร จะทำให้การเก็บรวบรวมข้อมูลหรือการคำนวณค่าสถิติต่างๆ สะดวกรวดเร็วยิ่งขึ้น จึงเป็นค่าที่น่าสนใจ นอกจากนี้ค่าอันดับยังช่วยแก้ปัญหาการเกิดข้อมูลผิดปกติ (robust relative to outlier) ได้ จึงมีนักวิจัยหลายท่านนำเสนอแผนการเลือกตัวอย่างที่ไม่ยุ่งยากซับซ้อน

และมีประสิทธิภาพโดยใช้ค่าอันดับแทนค่าวัดขนาดเมื่อค่าตัวแปร x เป็นค่าที่ไม่ทราบ และมีค่าความแปรปรวนของตัวประมาณค่ายอดรวมต่ำ กรณีตัวอย่างของการใช้ค่าอันดับของตัวแปรที่สนใจศึกษาซึ่งไม่ทราบค่าแต่สามารถจัดอันดับหน่วยในประชากรได้เกือบทุกหน่วย ได้แก่ การจ่ายกระแสไฟฟ้าของ 160 เครื่องของหน่วยงานรัฐ Tennessee (Tennessee Valley Authority , TVA) ดำเนินการจ่ายกระแสไฟฟ้าให้แก่ประชากร 2.5 ล้านครัวเรือน ครอบคลุม 8 รัฐ หน่วยวิจัยของ TVA ต้องการประมาณความต้องการกระแสไฟฟ้ารวมของประชากรในช่วงเวลาที่กำหนด โดยทำการลุ่มตัวอย่างแบบแบ่งกลุ่มหลายขั้น หน่วยตัวอย่างคือเครื่องจ่ายกระแสไฟฟ้า ทั้งหมด 160 เครื่อง แผนงานลุ่มตัวอย่างขั้นแรกจะทำการเลือกเครื่องจ่ายกระแสไฟฟ้า แบบการเลือกตัวอย่างเชิงความน่าจะเป็นที่เป็นสัดส่วนกับขนาด ทำให้ทราบความต้องการรวมของประชากรต่อเครื่องจ่ายกระแสไฟฟ้าในระยะเวลาหนึ่งเดือน ในกรณีนี้ถ้าค่าตัวแปร y_i เป็นความต้องการของประชากรต่อเครื่องจ่ายกระแสไฟฟ้าเครื่องที่ i ในช่วงเวลาหนึ่ง จะทำให้ทราบค่าอันดับของความต้องการกระแสไฟฟ้าสำหรับ 160 เครื่องอย่างสมเหตุสมผล

ในกรณีที่กลุ่มต่างๆ ถูกเลือกด้วยความน่าจะเป็นไม่เท่ากันแบบไม่ใส่คืนนั้น การหาความน่าจะเป็นที่กลุ่ม i ถูกเลือกเข้ามาในตัวอย่างจะยุ่งยากมากขึ้นกว่ากรณีที่เลือกแบบใส่คืน เนื่องจากความน่าจะเป็นที่กลุ่ม i ถูกเลือกในการเลือกครั้งที่ j จะขึ้นอยู่กับผลการเลือกที่ผ่านมาว่ามีกลุ่มใดถูกเลือกไปแล้วบ้างในการเลือกครั้งที่ 1 ถึงครั้งที่ (j-1) และความน่าจะเป็นที่กลุ่ม i จะอยู่ในตัวอย่างขนาด n ก็ไม่เท่ากับ np_i อีกต่อไป เพื่อให้การประมาณค่าเป็นไปได้ในทางปฏิบัติ ได้มีผู้คิดค้นวิธีการต่างๆหลายวิธี เช่น ดำเนินการโดยการเปลี่ยนรูปตัวประมาณแต่คงวิธีการเลือกตัวอย่างไว้ หรือปรับวิธีการเลือกตัวอย่างแต่คงรูปตัวประมาณไว้

ในการศึกษาครั้งนี้ ผู้วิจัยจะศึกษาว่าการจัดอันดับหน่วยในประชากรให้มีความสัมพันธ์อย่างสมเหตุสมผลกับตัวแปรที่สนใจที่ไม่ทราบค่าได้อย่างไร และ ค่าอันดับสามารถใช้แทนค่าวัดขนาดได้ดีเพียงไรเมื่อเทียบกับตัวแปรปริมาณหรือตัวแปร x ที่มีความสัมพันธ์กับ y ภายใต้แผนการเลือกตัวอย่างแบบ pps แบบไม่ใส่คืน ซึ่งเป็นแผนการเลือกตัวอย่างที่เสนอโดย Vasantha Kumar , E. ,Srivenkataramana , T. และ Srinath , K.P. และแผนการเลือกตัวอย่างที่เสนอโดย Tommy Wright ทั้งสองแผนการเลือกนี้เป็นแผนที่ปรับใหม่เพื่อลดความยุ่งยากซับซ้อนในการคำนวณค่าความแปรปรวนของตัวประมาณค่ายอดรวมประชากรและค่าความน่าจะเป็นที่หน่วยในประชากรจะถูกเลือกแบบไม่ใส่คืน เมื่อขนาดตัวอย่างมากกว่า 2 แต่สามารถคงรูปตัวประมาณค่ายอดรวมประชากรของ Horvitz- Thomsonได้

หอสมุดกลาง สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

1.2 ผลงานวิจัยที่เกี่ยวข้อง

งานวิจัยของ Vasantha Kumar , E. ,Srivenkataramana , T. และ Srinath , K.P.(1996) ได้เสนอแผนการเลือกตัวอย่างเชิงความน่าจะเป็นที่เป็นสัดส่วนกับขนาด โดยใช้ค่าอันดับเป็นค่าวัดขนาด เมื่อหน่วยไม่มีค่าซ้ำกัน และได้ทำการเปรียบเทียบประสิทธิภาพของแผนการเลือกตัวอย่างเชิงความน่าจะเป็นรวมที่เป็นสัดส่วนกับขนาด (inclusion probability proportional to size หรือ IPPS) ที่เสนอขึ้นใหม่ และ PPS แบบใส่คืนและไม่ใส่คืนโดยใช้ค่าตัวแปร x แทนค่าวัดขนาด เมื่อเลือกตัวอย่างขนาด 2 จากประชากรธรรมชาติ 5 ชุดที่มีค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปร x และ y อยู่ในช่วง 0.80 ถึง 0.99 โดยที่ IPPS(1) เป็นการสุ่มกสุ่มตัวอย่างแรกด้วยความน่าจะเป็นที่เป็นสัดส่วนกับค่าของตัวแปร x จากประชากรขนาด N และเลือกกลุ่มที่สองด้วยความน่าจะเป็นแบบมีเงื่อนไขกับการสุ่มครั้งแรก และเป็นสัดส่วนกับค่าอันดับที่จัดเรียงตามค่าตัวแปร x ของ N-1 หน่วยประชากร(กลุ่ม)ที่เหลือจากการสุ่มครั้งแรก และ IPPS(2) เป็นการสุ่มกสุ่มในประชากรแบบไม่ใส่คืน ด้วยความน่าจะเป็นที่เป็นสัดส่วนกับค่าอันดับทั้งสองครั้ง ดังตารางที่1.1 และ 1.2

ตารางที่ 1.1 แสดงลักษณะของประชากร

ที่	N	$\rho(x,y)$	X	y	Source
1	12	0.80	จำนวนคนรวม	จำนวนคนขาดงาน	Yates(1960) , p.159, Kraals 1-12.
2	10	0.85	จำนวนหลุมกล้วย	จำนวนเครือกล้วย	Sukhatme and Sukhatme (1970) , p.166, units 11-20.
3	10	0.87	พื้นที่เพาะปลูกรวม	พื้นที่ปลูกข้าว	Sukhatme and Sukhatme (1970) , p.51,villages 1-10.
4	14	0.93	พื้นที่ที่ใช้ปลูกข้าวโพดเป็นเอเคอร์ ปี 1958	พื้นที่ที่ใช้ปลูกข้าวโพดเป็นเอเคอร์ ปี 1960	Rao (1963) , pp. 202-215.
5	16	0.99	จำนวนประชากร ปี 1920	จำนวนประชากร ปี 1930	Cochran (1963), p. 156.

การเปรียบเทียบประสิทธิภาพ (relative efficiencies) ได้ทำการเปรียบเทียบเป็นเปอร์เซ็นต์กับแผนการเลือกตัวอย่าง PPSWR และให้ประสิทธิภาพของวิธีนี้เป็น 100%

ตารางที่ 1.2 แสดงค่าเปอร์เซนต์ของประสิทธิภาพ ของแผนการเลือกตัวอย่าง PPSWOR , IPPS แบบที่ 1 และ 2 เปรียบเทียบกับ PPSWR ซึ่งมีประสิทธิภาพเป็น 100% ของประชากรธรรมชาติ 5 ชุด ด้วยขนาดตัวอย่างเท่ากับ 2

population	PPSWR	PPSWOR	Cheme	
			IPPS (1)	IPPS (2)
1	100	110.83	110.90	69.65
2	100	109.51	111.29	7.19
3	100	122.87	116.35	40.70
4	100	117.45	110.10	12.62
5	100	99.99	104.46	9.26

จากประชากรธรรมชาติ 5 ชุดดังกล่าว ประสิทธิภาพของแผนการเลือกตัวอย่าง IPPS(1) ดีกว่าแผน PPSWR และดีกว่าแผน PPSWOR ด้วย แต่ IPPS(2) ซึ่งใช้ค่าอันดับที่จัดเรียงตามค่าตัวแปรxแทนค่าวัดขนาดจะมีประสิทธิภาพต่ำ ซึ่งจะเห็นว่าประสิทธิภาพของแผนการเลือกตัวอย่างแบบ PPSWR เป็น 100% ในขณะที่ประสิทธิภาพของ IPPS(2) อยู่ในช่วง 7% ถึง 70% เท่านั้น

ผลงานของ Hansen และ Herwitz (1943) แสดงให้เห็นว่า ตัวประมาณค่ายอดรวมประชากรที่ได้จากแผนการเลือกตัวอย่างเชิงความน่าจะเป็นที่เป็นสัดส่วนกับขนาด โดยใช้ค่าของข้อมูล x เป็นค่าวัดขนาดและมีความน่าจะเป็นที่กลุ่ม i จะถูกเลือกเป็นตัวอย่างสำหรับตัวอย่างขนาด n มีค่าเท่ากับ np_i (probability proportional to size(πps) sampling) เมื่อ p_i เป็นความน่าจะเป็นที่กลุ่ม i จะถูกเลือก จะมีประสิทธิภาพมากกว่าตัวประมาณที่ได้จากการเลือกตัวอย่างด้วยความน่าจะเป็นเท่ากัน ถึงแม้ว่าวิธี πps นี้จะเป็นวิธีที่สะดวกต่อการคำนวณ และสามารถหาตัวประมาณที่ไม่เอนเอียงของค่าพารามิเตอร์ และค่าประมาณที่ไม่เอนเอียงของค่าความแปรปรวนของตัวประมาณได้ แต่เป็นวิธีการเลือกแบบใส่คืน และเพื่อปรับปรุงแผนการเลือกตัวอย่างเป็นการเลือกแบบไม่ใส่คืน (πpswor) เพื่อให้กลุ่มตัวอย่างที่ถูกเลือกแล้วไม่สามารถถูกเลือกใหม่ได้อีก และทำ

ให้ได้ตัวประมาณที่มีประสิทธิภาพมากกว่าการเลือกแบบใส่คืน แต่การหาค่าความน่าจะเป็นที่กลุ่มประชากรจะถูกเลือกจะยุ่งยากและซับซ้อนขึ้น เมื่อขนาดตัวอย่างมากกว่า 2 Tommy Wright (1990) จึงเสนอแผนการเลือกตัวอย่างที่เลียนแบบวิธี πpswr แต่เป็นการเลือกแบบไม่ใส่คืนโดยจัดอันดับหน่วยประชากรจากน้อยไปมากตามค่าของตัวแปร x และแบ่งประชากรขนาด N เป็น n ชั้นภูมิ แต่ละชั้นภูมิมีขนาดเท่าๆกันเป็น 2k และมีลักษณะที่คล้ายคลึงกัน แล้วเลือกตัวอย่างจากแต่ละชั้นภูมิขนาด 1 ตัวอย่างอย่างเป็นอิสระกันทำให้สามารถหาตัวประมาณที่ไม่เอนเอียงของค่ารวมประชากรและค่าประมาณค่าความแปรปรวนของตัวประมาณ โดยประมาณจากค่าประมาณที่ไม่เอนเอียงของค่าความแปรปรวนของตัวประมาณ Horvitz-Thompson ภายใต้การเลือกตัวอย่างแบบใส่คืนได้ และเมื่อเปรียบเทียบ ค่าความแปรปรวนของตัวประมาณค่ายอดรวมประชากรที่ได้จากแผนการเลือกตัวอย่างของ Wright กับ แผนการเลือกตัวอย่าง πpswr โดยใช้ค่าอันดับที่จัดเรียงตามค่าของตัวแปร xแทนค่าวัดขนาด พบว่าจากประชากรตัวอย่าง5ประชากร เมื่อใช้แผนการเลือกตัวอย่างของ Wright มีค่าความแปรปรวนของตัวประมาณค่ายอดรวมต่ำกว่า แต่วิธีของ Wright อาจไม่เหมาะสมในกรณีที่ $N \neq 2kn$ และเมื่อสามารถจัดอันดับหน่วยในประชากรและให้ค่าอันดับที่มีความสัมพันธ์กับตัวแปรที่สนใจแล้ว การใช้ค่าอันดับเป็นค่าวัดขนาดภายใต้แผนการเลือกตัวอย่าง πpswor ของ Wright จะทำให้ได้ตัวประมาณค่ายอดรวมประชากรที่มีประสิทธิภาพดีกว่า การใช้ค่าตัวแปร x แทนค่าวัดขนาดภายใต้แผนการเลือกตัวอย่าง πpswr ในกรณีที่ x มีความสัมพันธ์กับ y น้อยกว่าความสัมพันธ์ของ y และค่าอันดับของหน่วยประชากร

1.3 วัตถุประสงค์ของการวิจัย

1. เพื่อเปรียบเทียบประสิทธิภาพของตัวประมาณค่ารวมประชากร ภายใต้แผนการเลือกตัวอย่างเชิงความน่าจะเป็นที่เป็นสัดส่วนกับขนาด แบบไม่ใส่คืน กรณีที่ใช้ค่าอันดับและกรณีที่ใช้ค่าของตัวแปรที่มีความสัมพันธ์กับตัวแปรที่สนใจ เป็นค่าวัดขนาด

2. เพื่อเปรียบเทียบวิธีการให้ค่าอันดับกับหน่วยประชากร ภายใต้แผนการเลือกตัวอย่างเชิงความน่าจะเป็นที่เป็นสัดส่วนกับขนาด แบบไม่ใส่คืน ในกรณีที่ใช้ค่าอันดับแทนค่าวัดขนาด

1.4 สมมติฐานของการวิจัย

ตัวประมาณค่ารวมประชากร ภายใต้การเลือกตัวอย่างpps แบบไม่ใส่คืน โดยที่ความสัมพันธ์ระหว่างตัวแปรที่สนใจ (y) กับตัวแปร (x) ที่มีความสัมพันธ์กับตัวแปรที่สนใจมีค่าอยู่ระหว่าง 0.50 - 0.59 และขนาดตัวอย่างไม่เกิน 9 กรณีใช้ค่าอันดับเป็นค่าวัดขนาดมีประสิทธิภาพเท่าเทียมกับกรณีใช้ค่าของตัวแปรที่มีความสัมพันธ์กับตัวแปรที่สนใจ เป็นค่าวัดขนาด

1.5 ขอบเขตของการวิจัย

ศึกษาประสิทธิภาพของตัวประมาณค่ายอดรวมประชากรโดยอาศัยข้อมูลทุติยภูมิ เมื่อกำหนด

1. ความสัมพันธ์ระหว่างตัวแปรที่สนใจ (y) กับตัวแปร (x) ที่มีความสัมพันธ์กับตัวแปรสนใจมีค่าระหว่าง 0.50 และ 0.99
2. ขนาดตัวอย่าง 2, 3, 4, 6 และ 9
3. ค่าวัดขนาด เป็น
   3.1 ค่าของตัวแปร (x) ที่มีความสัมพันธ์กับตัวแปรที่สนใจ (y)
   3.2 ค่าอันดับของหน่วยประชากรที่จัดตามค่าตัวแปร x
   3.3 ค่าอันดับที่ปรับ เพื่อลดค่าความแตกต่างระหว่างค่าอันดับกับค่าของตัวแปร x ในเชิงปริมาณ
4. แผนการเลือกตัวอย่าง
   4.1 แผนการเลือกตัวอย่าง pps แบบไม่ใส่คืนของ Vasantha kumar E. ,Srivenkatamana,T.และ Srinath ,K.P. (1996)
   4.2 แผนการเลือกตัวอย่าง pps แบบไม่ใส่คืน ของ Wright (1990)

โดยมีวิธีการกำหนดขอบเขตของชั้นภูมิ คือ

4.2.1 วิธีของ Tommy Wright (1990)

4.2.2 วิธีค่าสะสมของค่ารากที่สองของความถี่ ($\sqrt{f}$) โดยอาศัยค่าของตัวแปรx

1.6 ข้อกำหนดของการวิจัย

<u>กรณีประชากรขนาดเล็กและสามารถหาตัวประมาณของทุกชุดตัวอย่างได้</u>

ค่าวัดความถูกต้องของตัวประมาณค่ารวมประชากร คือ ค่าเฉลี่ยของเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์ (mean absolute persent error)

$$MAPE = \frac{100}{m} * \sum_{i=1}^{m} \left| \frac{\hat{Y}_i - Y}{Y} \right|$$ เมื่อ m เป็นจำนวนชุดตัวอย่างที่เป็นไปได้

<u>กรณีประชากรขนาดใหญ่และมีจำนวนชุดตัวอย่างที่เป็นไปได้จำนวนมาก</u>

เลือกตัวอย่างมากกว่า 1 ชุดตัวอย่าง ภายใต้ข้อกำหนด

1. การแจกแจงของตัวประมาณค่ายอดรวมประชากรจากแต่ละชุดตัวอย่างที่เป็นไปได้ มีการแจกแจงแบบปกติ
2. ค่าเฉลี่ยตัวประมาณค่ายอดรวมประชากรจากชุดตัวอย่างที่สุ่มมาได้ มีค่าเท่ากับค่ายอดรวมประชากรหรือค่าจริง
3. ค่าความแปรปรวนของตัวประมาณค่ายอดรวมประชากรจากชุดตัวอย่างที่สุ่มมาต้องมีค่าต่ำกว่าค่าความแปรปรวนของตัวประมาณค่ายอดรวมประชากรของทุกชุดตัวอย่างที่เป็นไปได้

คำนวณหาค่า APE ของแต่ละชุดตัวอย่างที่สุ่มมาได้ แล้วหาค่าเฉลี่ยของ APE จะได้ค่าประมาณของ MAPE

พิจารณาความแตกต่างของค่าเฉลี่ยของ MAPE ที่ระดับนัยสำคัญ 0.05 จากปัจจัยที่มีผลกระทบต่อค่าเฉลี่ย MAPE ในที่นี้ได้แก่ แผนการเลือกตัวอย่าง , ระดับความสัมพันธ์ระหว่างตัวแปร x และ y , ตัวแปรที่ใช้แทนค่าวัดขนาด และ ขนาดตัวอย่าง ภายใต้ขอบเขตการวิจัยดังกล่าวข้างต้น

1.7 ประโยชน์ของการวิจัย

เป็นแนวทางในการให้ค่าอันดับกับหน่วยประชากร เพื่อใช้เป็นค่าวัดขนาด ได้อย่างมีประสิทธิภาพเท่าเทียมกับการใช้ค่าของตัวแปรที่มีความสัมพันธ์อย่างสูงกับตัวแปรที่สนใจ