บทที่ 1

บทนำ

1.1 ความสำคัญและความเป็นมาของปัญหา

ในการสำรวจตัวอย่างด้วยวิธีการสัมภาษณ์มักมีความคลาดเคลื่อนที่ไม่ได้เกิดจากการเลือกตัวอย่าง (non-sampling error) ที่สำคัญ 2 ประการ คือ ประการแรกผู้ถูกสัมภาษณ์ปฏิเสธที่จะตอบคำถาม และประการที่สอง ผู้ถูกสัมภาษณ์ตอบคำถามไม่ตรงตามความจริง ความคลาดเคลื่อนทั้งสองนี้จะยิ่งเพิ่มขึ้นถ้าคำถามที่สัมภาษณ์เป็นเรื่องที่เกี่ยวกับการกระทำที่ผิดกฎหมาย หรือพฤติกรรมที่เบี่ยงเบนไปจากการยอมรับของสังคม เช่น การติดยาเสพติด การทำแท้ง หรือการมีเพศสัมพันธ์ก่อนสมรส ทั้งนี้เพราะข้อเท็จจริงของเรื่องเหล่านี้ถือว่าเป็นความลับส่วนบุคคลซึ่งทุกคนพยายามที่จะปกปิด โดยเฉพาะอย่างยิ่งบุคคลที่เคยทำผิดกฎหมาย หรือเคยมีพฤติกรรมดังที่กล่าวมาแล้วข้างต้น ดังนั้นข้อมูลที่เก็บรวบรวมได้จึงขาดความถูกต้อง และทำให้เกิดความเอนเอียง (bias) ในการประมาณค่าพารามิเตอร์ของประชากร

วอร์เนอร์ (1965:63-69) ได้คิดวิธีการแก้ปัญหาความคลาดเคลื่อนทั้งสองขึ้น เรียกว่า เทคนิคการเลือกตอบคำถามอย่างสุ่ม (Randomized Response Technique :RRT) กล่าวคือแทนที่จะให้ผู้ถูกสัมภาษณ์ตอบคำถามของเรื่องปกปิดที่ต้องการศึกษาโดยตรง ก็ให้ผู้ถูกสัมภาษณ์สุ่มคำถาม 1 คำถามจากคำถาม 2 คำถามที่เกี่ยวข้องกัน (related questions) โดยไม่บอกให้ผู้สัมภาษณ์ทราบว่าสุ่มได้คำถามใด ซึ่งคำถามแต่ละคำถามได้ถูกกำหนดความน่าจะเป็นที่จะถูกสุ่มไว้ล่วงหน้าแล้วโดยผู้สำรวจ สำหรับคำถามที่เกี่ยวข้องกันนั้น คำถามแรก เป็นคำถามของเรื่องปกปิดที่ต้องการศึกษา เช่น "คุณเคยทำแท้ง ใช่หรือไม่" คำถามที่สอง เป็นนิเสธ (negative) ของคำถามแรก เช่น "คุณไม่เคยทำแท้ง ใช่หรือไม่" และคำตอบของผู้ถูกสัมภาษณ์ในแต่ละคำถามที่สุ่มได้จะต้องไม่สื่อความหมายพอที่จะทำให้ผู้สัมภาษณ์ทราบว่าเป็นคำตอบของคำถามใด เช่น "ใช่" (yes) กับ "ไม่ใช่" (no) หรือ "เคย" กับ "ไม่เคย" จะเห็นว่าวิธีการนี้ทำให้ความลับของผู้ที่มีลักษณะปกปิดก็ยังคงเป็นความลับต่อไป เมื่อผู้ถูกสัมภาษณ์ทุกคนเข้าใจในวิธีการและตอบคำถามตามความจริงแล้ว ข้อมูลทั้งหมดสามารถนำไปประมาณค่าสัดส่วนของประชากรที่มีลักษณะปกปิด (π) ได้โดยใช้หลักวิชาการทางสถิติ

ในการประมาณค่าสัดส่วนของประชากรที่มีลักษณะปกปิดด้วยวิธีการของวอร์เนอร์นั้น คำถามที่ให้ผู้ถูกสัมภาษณ์สุ่มขึ้นมาตอบจะเกี่ยวข้องกับเรื่องปกปิดทั้งสองข้อ ซึ่งอาจทำให้ผู้ถูกสัมภาษณ์ไม่สบายใจที่จะตอบคำถาม และอาจตัดสินใจไม่ให้ความร่วมมือได้ แต่ถ้าเปลี่ยนคำถามที่ใช้เป็นคำถามที่ไม่เกี่ยวข้องกัน (unrelated questions) คือ คำถามแรกยังคงเป็นคำถามของเรื่องปกปิดเหมือนเดิม ส่วนคำถามที่สองนั้นเป็นคำถามของเรื่องทั่วๆ ไปที่ไม่จำเป็นต้องปกปิด เช่น "คุณเกิดเดือนสิงหาคม ใช่หรือไม่" หรือ "คุณเคยไปจังหวัดเชียงใหม่หรือไม่" กรีนเบอร์ก และคณะ (1969:520-539) พบว่าความร่วมมือในการสัมภาษณ์ และประสิทธิภาพของตัวประมาณจะเพิ่มขึ้น ถ้าคำถามที่สองให้ค่าสัดส่วนของประชากรที่มีลักษณะทั่วไป (π_u) สอดคล้องกับสัดส่วนของประชากรที่มีลักษณะปกปิด กล่าวคือถ้าสัดส่วนของประชากรที่มีลักษณะปกปิดมีค่าสูงหรือต่ำแล้ว ค่าสัดส่วนของประชากรในคำถามที่สองจะต้องสูงหรือต่ำตามไปด้วย แต่ตามปกติแล้ว ผู้สำรวจมักจะไม่ทราบค่า π ดังนั้นจึงเกิดปัญหาการเลือกคำถามที่ให้ค่า π_u ที่สอดคล้องกับ π กรณีเช่นนี้ แลนเก้ (1975:80-83) ได้แนะนำให้ใช้คำถามที่มีค่า π_u สูงๆ ทั้งนี้เพราะ π_u สูงๆ จะทำให้ผู้ถูกสัมภาษณ์ตอบ "ใช่" มากขึ้น และนอกจากนั้นยังทำให้ผู้ที่มีลักษณะปกปิดสบายใจที่จะตอบ "ใช่" มากขึ้นด้วย แต่ในทางปฏิบัติ บางครั้งผู้สำรวจไม่สามารถหาคำถามที่ให้ค่า π_u สูงๆ ได้ การแก้ปัญหานี้วิธีของสรชัย (2524:1-23) สามารถช่วยได้โดยใช้เครื่องมือสุ่มเลือก (randomizing device) เป็นตัวกำหนดค่า π_u ซึ่งมีวิธีการดังนี้ สมมติเครื่องมือสุ่มเลือกที่ใช้เป็น ไพ่ 1 ชุดประกอบด้วยไพ่สีแดง และไพ่สีดำในอัตราส่วนที่ผู้สำรวจกำหนดขึ้นในการสัมภาษณ์ ผู้ถูกสัมภาษณ์จะสับไพ่จนเข้ากันดีแล้วดึงไพ่ 1 ใบ โดยไม่ให้ผู้สัมภาษณ์ทราบว่าดึงได้ไพ่สีอะไร ถ้าดึงได้ไพ่สีแดงให้ตอบคำถามของเรื่องปกปิด เช่น "คุณเคยทำแท้ง ใช่หรือไม่" แต่ถ้าดึงได้ไพ่สีดำให้ตอบว่า "ใช่" เสมอ ซึ่งเหมือนกับถามว่า "ไพ่ที่ท่านดึงได้เป็นไพ่สีดำ ใช่หรือไม่" จะเห็นว่าวิธีการนี้ก็คือรูปแบบหนึ่งของการใช้คำถามที่ไม่เกี่ยวข้องกันโดยกำหนดสัดส่วนของประชากรในคำถามที่สองให้เท่ากับ 1 ถึงแม้ว่าวิธีของสรชัยจะสามารถแก้ปัญหาดังกล่าวได้ แต่ถ้าพิจารณาให้ดีแล้ววิธีการนี้อาจทำให้ผู้ถูกสัมภาษณ์เกิดความไม่ไว้วางใจได้เพราะ ถ้าเขาตอบว่า "ใช่" ผู้สัมภาษณ์จะไม่ทราบว่า "ใช่" เนื่องจากเคยทำแท้ง หรือเนื่องจากไพ่ที่สุ่มได้เป็นไพ่สีดำ ถ้าเขาตอบว่า "ไม่ใช่" ผู้สัมภาษณ์จะทราบทันทีว่า ผู้ตอบไม่เคยทำแท้ง ทั้งนี้เพราะ คำตอบ "ไม่" จะมีเฉพาะในคำถามแรกเท่านั้นส่วนคำถามที่สองไม่มี ดังนั้นจะเห็นว่าคำตอบของผู้ถูกสัมภาษณ์สื่อความหมายพอที่ผู้สัมภาษณ์จะทราบว่าเป็นคำตอบของคำถามใด ซึ่งไม่ตรงกับวัตถุประสงค์ของเทคนิคการเลือกตอบคำถามอย่างสุ่ม จุดนี้เองที่ทำให้ผู้วิจัยสนใจจะศึกษา และต้องการที่จะพัฒนาวิธีของสรชัยให้ดีขึ้น เพื่อประโยชน์ในการประมาณค่าสัดส่วนของประชากร ที่มีลักษณะปกปิดต่อไป

1.2 วัตถุประสงค์ของการวิจัย

เพื่อเปรียบเทียบวิธีการประมาณค่าสัดส่วนของประชากรที่มีลักษณะปกปิด 3 วิธี

วิธีที่ 1 วิธีของวอร์เนอร์ (Warner's Method หรือ Method 1)
วิธีที่ 2 วิธีของสรชัย (Sorachai's Method หรือ Method 2)
วิธีที่ 3 วิธีที่พัฒนาจากวิธีที่ 2 (The Modified Sorachai's Method หรือ Method 3)

ทั้งในกรณีที่ผู้ถูกสัมภาษณ์ทุกคนตอบตามความจริง (completely truthful) และกรณีที่บางส่วนของผู้ถูกสัมภาษณ์ที่มีลักษณะปกปิดตอบไม่ตรงตามความจริง (less than completely truthful)

1.3 สมมติฐานของการวิจัย

วิธีการประมาณค่าสัดส่วนของประชากรที่มีลักษณะปกปิด วิธีที่ 3 มีประสิทธิภาพดีกว่าวิธีที่ 1 และ วิธีที่ 2 ทั้งในกรณีที่ผู้ถูกสัมภาษณ์ทุกคนตอบตามความจริง และกรณีที่บางส่วนของผู้ถูกสัมภาษณ์ที่มีลักษณะปกปิดตอบไม่ตรงตามความจริง

1.4 ข้อตกลงเบื้องต้น

1.4.1 ผู้ถูกสัมภาษณ์ทุกคนยินดีให้ความร่วมมือในการสัมภาษณ์

1.4.2 กรณีที่ผู้ถูกสัมภาษณ์ทุกคนตอบตามความจริงนั้น เป็นเพราะเขาเชื่อมั่นว่าวิธีการทั้ง 3 นี้ จะไม่สามารถเปิดเผยข้อเท็จจริงให้ผู้สัมภาษณ์ทราบได้

1.4.3 กรณีที่บางส่วนของผู้ถูกสัมภาษณ์ที่มีลักษณะปกปิดตอบไม่ตรงตามความจริงนั้นในแต่ละวิธีนั้น เป็นเพราะต้องการที่จะปกปิดข้อเท็จจริงของตนเท่านั้น ไม่ใช่เกิดจากความสับสนในวิธีการตอบคำถามของแต่ละวิธี

1.4.4 เกณฑ์ในการเลือกตัวประมาณสำหรับการวิจัยครั้งนี้แบ่งออกเป็น 2 กรณี

1.4.4.1 กรณีที่ผู้ถูกสัมภาษณ์ทุกคนตอบตามความจริง ตัวประมาณสัดส่วนของประชากรที่มีลักษณะปกปิดทั้ง 3 วิธี เป็นตัวประมาณที่ไม่เอนเอียง (unbiased estimator) ดังนั้น จะใช้ความแปรปรวนเป็นเกณฑ์ในการเลือกวิธีการประมาณค่าสัดส่วนที่มีลักษณะปกปิด

1.4.4.2 กรณีที่บางส่วนของผู้ถูกสัมภาษณ์ที่มีลักษณะปกปิดตอบไม่ตรงตามความจริง ตัวประมาณสัดส่วนของประชากรที่มีลักษณะปกปิดทั้ง 3 วิธี เป็นตัวประมาณที่เอนเอียง (biased estimator) ดังนั้น จะใช้ความคลาดเคลื่อนกำลังสองเฉลี่ยเป็นเกณฑ์ในการเลือกวิธีการประมาณค่าสัดส่วนที่มีลักษณะปกปิด

1.5 ขอบเขตของการวิจัย

1.5.1 กรณีที่ผู้ถูกสัมภาษณ์ทุกคนตอบตามความจริง การเปรียบเทียบประสิทธิภาพของตัวประมาณนั้นศึกษาได้ด้วยการพิสูจน์ โดยกำหนดค่าพารามิเตอร์ต่างๆ ดังต่อไปนี้

1.5.1.1 กำหนดสัดส่วนของประชากรที่มีลักษณะปกปิด (π) มีค่า [0,1]

1.5.1.2 ความน่าจะเป็นที่ผู้ถูกสัมภาษณ์จะสุ่มได้คำถามที่ปกปิด (p) มีค่า (0,1)

1.5.2 กรณีที่บางส่วนของผู้ถูกสัมภาษณ์ที่มีลักษณะปกปิดตอบไม่ตรงตามความจริงนั้น การเปรียบเทียบประสิทธิภาพของตัวประมาณนั้นทำโดยการทดลอง ดังนั้น จึงกำหนดค่าพารามิเตอร์ต่างๆ ดังต่อไปนี้

1.5.2.1 สัดส่วนของประชากรที่มีลักษณะปกปิดมีค่า 1 0.9 0.8 0.7 0.6 0.5 0.4 0.3 0.2 0.1 0.05 และ 0.01

1.5.2.2 ความน่าจะเป็นที่ผู้ถูกสัมภาษณ์จะสุ่มได้คำถามที่ปกปิดมีค่า 0.9 0.8 0.7 0.6 0.5 0.4 0.3 0.2 และ 0.1

1.5.2.3 ขนาดของกลุ่มตัวอย่างที่ใช้ในการวิจัยครั้งนี้เท่ากับ 100, 500 และ 1000

1.5.2.4 การประมาณค่าสัดส่วนของประชากรที่มีลักษณะปกปิดทั้ง 3 วิธี ที่นำมาศึกษาในครั้งนี้ ผู้สัมภาษณ์สามารถแสดง และอธิบายให้ผู้ถูกสัมภาษณ์เข้าใจ ว่าวิธีทั้ง 3 ไม่สามารถเปิดเผยข้อเท็จจริงของผู้ถูกสัมภาษณ์ได้ แต่อาจจะมีผู้ถูกสัมภาษณ์ที่มีลักษณะปกปิดเพียงบางส่วนเท่านั้นที่ยังไม่ไว้วางใจ จึงตอบคำถามไม่ตรงตามความจริง ดังนั้น ในการวิจัยครั้งนี้ จะกำหนดความน่าจะเป็นที่ผู้ถูกสัมภาษณ์ที่มีลักษณะปกปิดจะตอบคำถามของเรื่องปกปิดตามความจริง เป็น 1 0.95 0.90 0.70 และ 0.50

1.6 คำจำกัดความ

1.6.1 ประชากรที่มีลักษณะปกปิด หมายถึง ประชากรที่กระทำ/เคยทำผิดกฎหมาย หรือมี/เคยมีพฤติกรรมที่เบี่ยงเบนไปจากการยอมรับของสังคม เช่น การติดยาเสพติด การทำแท้ง การก่ออาชญากรรมต่างๆ หรือการมีเพศสัมพันธ์ก่อนการสมรส เป็นต้น

1.6.2 สัดส่วนของประชากรที่มีลักษณะปกปิด หมายถึง จำนวนประชากรที่มีลักษณะปกปิดต่อประชากรทั้งหมดที่ศึกษา

1.6.3 เครื่องมือสุ่มเลือก (randomizing device) หมายถึงเครื่องมือที่ใช้ในเทคนิคการเลือกตอบคำถามอย่างสุ่ม (Randomized Response Technique) เช่น ไพ่ 1 ชุด ประกอบด้วยไพ่สีแดง และไพ่สีดำ ในอัตราส่วนที่ผู้สำรวจกำหนดไว้ล่วงหน้า หรือกล่องพลาสติกสี่เหลี่ยมทึบทุกด้าน ยกเว้นด้านหน้าเพียงด้านเดียวที่ทำด้วยพลาสติกใสที่กึ่งกลางของด้านที่ทำด้วยพลาสติกทึบ และไม่ได้อยู่ตรงข้ามกับด้านที่เป็นพลาสติกใส เจาะรูเป็นวงกลมให้มีขนาดพอที่ลูกบอลพลาสติกขนาดเล็กเพียงลูกเดียวสามารถลอดออกมาค้างอยู่ในกล่องสี่เหลี่ยมขนาดเล็กที่รองรับลูกบอล ซึ่งลอดออกมาจากกล่องสี่เหลี่ยมที่บรรจุด้วยลูกบอลสีแดง และสีดำ ในอัตราส่วนที่ผู้สำรวจกำหนดไว้ล่วงหน้า สำหรับกล่องสี่เหลี่ยมที่ใช้รองรับลูกบอลเพียงลูกเดียวนี้ จะมีด้านหนึ่งที่ทำด้วยพลาสติกใส ซึ่งอยู่ด้านเดียวกับด้านพลาสติกใสของกล่องที่บรรจุลูกบอล

1.6.4 ประสิทธิภาพสัมพัทธ์ (Relative Efficiency) หรือ RE ของตัวประมาณ $\hat{\pi}_1$ เมื่อเทียบกับ $\hat{\pi}_2$ เป็นการเปรียบเทียบค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณ 2 ตัว ที่ประมาณค่าพารามิเตอร์เดียวกัน ในรูปอัตราส่วนของค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณหนึ่ง ต่อค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของอีกตัวหนึ่ง หรืออาจจะพิจารณาจากสูตรได้ดังนี้

$$RE(\hat{\pi}_1, \hat{\pi}_2) = \frac{MSE(\hat{\pi}_2)}{MSE(\hat{\pi}_1)}$$

1.7 ประโยชน์ของการวิจัย

เพื่อช่วยให้ผู้ที่ทำการสำรวจข้อมูลที่มีลักษณะปกปิด ได้เลือกใช้วิธีการประมาณค่าสัดส่วนของประชากรที่มีลักะปกปิดได้อย่างเหมาะสม และมีประสิทธิภาพ

หอสมุดกลาง สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย