

การประมาณค่าสัดส่วนของประชากรที่มีลักษณะปกปิด
โดยใช้เทคนิคการเลือกตอบคำถามอย่างสุ่ม
**Estimation of Population Proportion with Sensitive Characteristic
Using Randomized Response Technique**

ดลชาติ ตันติวานิช

Dolachart Tantivanich

สาขาวิชาสถิติ คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ถนนฉลองกรุง เขตลาดกระบัง กรุงเทพฯ 10520

บทคัดย่อ

เทคนิคการเลือกตอบคำถามอย่างสุ่มเป็นการนำทฤษฎีความน่าจะเป็นมาประยุกต์ใช้ในการสำรวจข้อมูลที่มีลักษณะปกปิด (sensitive) เพื่อเพิ่มความร่วมมือในการตอบคำถามปกปิดโดยที่ความลับของผู้ถูกสัมภาษณ์ไม่ถูกเปิดเผย บทความนี้มีวัตถุประสงค์เพื่อแสดงให้เห็นว่าเทคนิคนี้สามารถประมาณค่าสัดส่วนของประชากรที่มีลักษณะปกปิดได้อย่างไร

คำสำคัญ : เทคนิคการเลือกตอบคำถามอย่างสุ่ม ข้อมูลที่มีลักษณะปกปิด

Abstract

Randomized response technique applies probability theory to surveys dealing with sensitive characteristics in order to increase respondent cooperation in answering sensitive questions without revealing confidential information of respondents. The purpose of this article is to show how to use randomized response technique to estimate population proportion with sensitive characteristics.

Keywords : Randomized response technique, Sensitive characteristics.

1. บทนำ

การประมาณค่าสัดส่วนของประชากรที่มีลักษณะปกปิด (sensitive) ด้วยวิธีการสัมภาษณ์มักจะพบว่า ผู้ถูกสัมภาษณ์ส่วนใหญ่ไม่ค่อยให้ความร่วมมือในการตอบคำถาม หรือตอบคำถามไม่ตรงตามความเป็นจริง Warner [1] ได้แก้ปัญหานี้โดยใช้เทคนิคการเลือกตอบคำถามอย่างสุ่ม (Randomized Response Technique : RRT) เทคนิคนี้แทนที่จะให้ผู้ถูกสัมภาษณ์ตอบคำถามที่มีลักษณะปกปิดโดยตรง แต่จะให้ผู้ถูกสัมภาษณ์เลือกคำถาม 1 ข้อจากคำถาม 2 ข้อ ที่เกี่ยวข้องกัน (related questions) โดยใช้เครื่องมือสุ่มเลือก (Randomized Device) เช่น ไฟ 1 สำหรับ ซึ่งประกอบด้วย ไฟสีแดง และสีดำในอัตราส่วนที่กำหนดไว้ล่วงหน้า คำถามที่เกี่ยวข้องกันนั้น คำถามแรกจะเป็นคำถามของเรื่องปกปิดที่ต้องการศึกษา เช่น “คุณเคยทุจริตในการสอบใช่หรือไม่” ส่วนคำถามที่สองจะเป็นนิเสธของคำถามแรก คือ “คุณไม่เคยทุจริตในการสอบใช่หรือไม่” เป็นต้น วิธีการตอบคำถามให้ผู้ถูกสัมภาษณ์สุ่มไฟขึ้นมา 1 ใบ โดยไม่ให้ผู้สัมภาษณ์ทราบว่า เป็นไฟสีอะไร ถ้าสุ่มได้ไฟสีแดงให้ตอบคำถามแรกว่า “ใช่หรือไม่” และถ้าสุ่มได้ไฟสีดำ ให้ตอบคำถามที่สองว่า “ใช่หรือไม่” การตอบ “ใช่” กับ “ไม่ใช่” ไม่สามารถชี้ว่าผู้ถูกสัมภาษณ์เคยทุจริต หรือไม่เคยทุจริต ดังนั้นเมื่อผู้ถูกสัมภาษณ์เข้าใจในวิธีการดังกล่าวและเชื่อมั่นว่าความลับของตนจะไม่ถูกเปิดเผย และตอบคำถามตามความเป็นจริงแล้ว ผลลัพธ์ที่ได้สามารถนำไปประมาณค่าสัดส่วนของประชากรที่มีลักษณะปกปิดได้ โดยใช้ทฤษฎีความน่าจะเป็น

Greenberg *et al.* [2] ได้พัฒนาเทคนิคนี้โดยใช้คำถามที่ไม่เกี่ยวข้องกัน คือ คำถามแรกยังคงเป็นคำถามของเรื่องปกปิด เช่น “คุณเคยทุจริตในการสอบใช่หรือไม่” ส่วนคำถามที่สองเป็นคำถามอื่น ๆ ที่ไม่เกี่ยวข้องกับเรื่องปกปิด เช่น “คุณเคยไปเที่ยวเชียงใหม่ใช่หรือไม่” จากผลการวิจัย พบว่าผู้ถูกสัมภาษณ์ให้ความร่วมมือมากขึ้น นอกจากนั้นยังมีผู้ที่พัฒนาเทคนิคดังกล่าวเพื่อประยุกต์ใช้กับการสำรวจในชีวิตจริง เช่น Lanke [3], Chaudhuri and Mukerjee [4], Mangat and Singh [5] และ Kuk [6] เป็นต้น สำหรับประเทศไทยมีผู้ที่สนใจศึกษาเทคนิคนี้หลายคน เช่น สรชัย [7] ดลชาติ [8] นุชนาด [9] Noodchanath and Surin [10] สุวรรณภา และสุรินทร์ [11] ดลชาติ [12-13] ดลชาติ และ พรชัย [14] เป็นต้น

หลังจาก ดลชาติ และ พรชัย [14] ได้เผยแพร่เทคนิคการเลือกตอบคำถามอย่างสุ่มลงในวารสารวิทยาศาสตร์ ลาดกระบังแล้ว หลายคนสงสัยว่าเทคนิคดังกล่าวจะสามารถประมาณค่าสัดส่วนของประชากรที่มีลักษณะปกปิดได้อย่างไร จึงเป็นที่มาของการเขียนบทความนี้

2. การประมาณค่าสัดส่วนของประชากรที่มีลักษณะปกปิดโดยใช้เทคนิคการเลือกตอบ

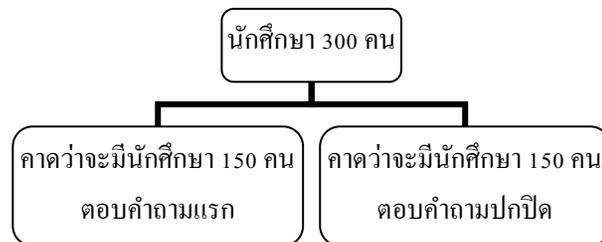
คำถามอย่างสุ่ม

ตัวอย่างต่อไปนี้เป็นแนวคิดของการใช้เทคนิคการเลือกตอบคำถามอย่างสุ่มเพื่อประมาณค่าสัดส่วนของนักศึกษาที่เลขทศวรรษในการสอบ โดยใช้ลูกเต๋าเป็นเครื่องมือสุ่มเลือก สุ่มตัวอย่าง นักศึกษาจำนวน 300 คน ให้นักศึกษาแต่ละคนโยนลูกเต๋าสี่เหลี่ยมที่สมดุล 1 ลูก 1 ครั้ง โดยไม่ให้ผู้วิจัย ทราบว่าลูกเต๋ารับขึ้นหน้าอะไร ถ้าลูกเต๋ารับขึ้นหน้า 1, 2 และ 3 ให้ตอบคำถามแรกว่า “ท่านเกิดในหก เดือนแรกของปี (มกราคม-มิถุนายน) ใช่หรือไม่” แต่ถ้าลูกเต๋ารับขึ้นหน้า 4, 5 และ 6 ให้ตอบคำถาม ปกปิด คือ “ท่านเลขทศวรรษในการสอบ ใช่หรือไม่”

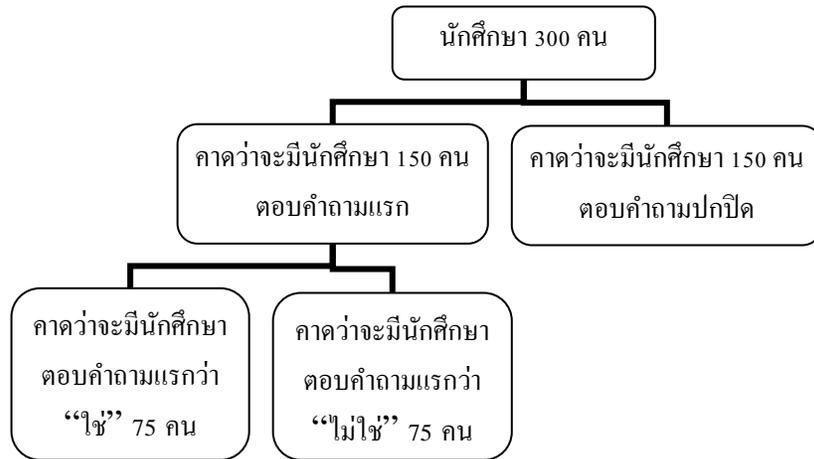
ในการใช้เทคนิคนี้ ก่อนการตอบคำถาม ผู้วิจัยจะต้องอธิบายให้นักศึกษาเข้าใจว่าวิธีการ ดังกล่าวจะไม่ทำให้ความลับของตนเองถูกเปิดเผย เพราะ ถ้านักศึกษาตอบว่า “ใช่” ผู้วิจัยจะไม่ ทราบว่า “ใช่” เนื่องจากเขาเกิดในหกเดือนแรกของปี หรือ เลขทศวรรษในการสอบ เช่นเดียวกับ คำตอบว่า “ไม่ใช่” ผู้วิจัยก็ไม่ทราบว่าเป็น “ไม่ใช่” เนื่องจากเขาไม่ได้เกิดในหกเดือนแรกของปี หรือ ไม่เลขทศวรรษในการสอบ ทั้งนี้เพราะผู้วิจัยไม่ทราบว่าลูกเต๋านักศึกษาโยนขึ้นหน้าอะไร เมื่อนักศึกษาเชื่อมั่นว่าวิธีการนี้ไม่ทำให้ความลับของตนเองถูกเปิดเผย และให้ความร่วมมือที่จะตอบคำถาม ตามความเป็นจริง คำตอบ “ใช่” กับ “ไม่ใช่” ของนักศึกษา 300 คน สามารถนำไปประมาณค่า สัดส่วนของนักศึกษาที่เลขทศวรรษ โดยใช้ทฤษฎีความน่าจะเป็นเข้ามาประยุกต์ได้ดังนี้ สมมติว่า ในการศึกษาคั้งนี้มีนักศึกษาตอบ “ใช่” จากทั้งสองคำถาม จำนวน 130 คน

1. ความน่าจะเป็นที่ลูกเต๋ารับขึ้นหน้า 1, 2 และ 3 เท่ากับ $\frac{1}{2}$ เพราะฉะนั้น เราคาดว่า จะมี

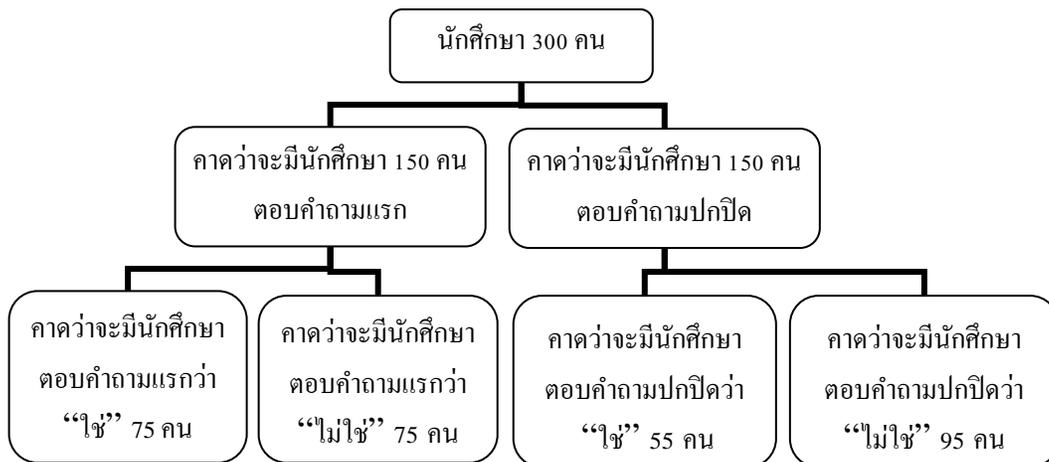
นักศึกษาจำนวน $\frac{1}{2} \times 300 = 150$ คน ที่ตอบคำถามแรก คือ “ท่านเกิดในหกเดือนแรกของปี ใช่หรือไม่” และ อีก 150 คน ตอบคำถามปกปิด คือ “ท่านเลขทศวรรษในการสอบ ใช่หรือไม่” ดังรูปต่อไปนี้



2. ความน่าจะเป็นที่นักศึกษาจะเกิดในหกเดือนแรกของปีเท่ากับ $\frac{1}{2}$ เพราะฉะนั้น จาก นักศึกษาที่คาดว่าจะตอบคำถามแรกจำนวน 150 คน เราคาดว่าจะมีนักศึกษาที่เกิดในหกเดือนแรกของปีเท่ากับ $\frac{1}{2} \times 150 = 75$ คน นั่นคือ มีนักศึกษาที่ตอบคำถามแรกว่า “ใช่” จำนวน 75 คน และ ตอบว่า “ไม่ใช่” จำนวน $150 - 75 = 75$ คน ดังรูปต่อไปนี้



3. จากนักศึกษาทั้งหมด 300 คน มีนักศึกษาที่ตอบคำถามทั้งสองข้อว่า “ใช่” 130 คน และนักศึกษาที่ตอบว่า “ใช่” สำหรับคำถามแรก 75 คน เพราะฉะนั้น เราคาดว่าจะมีนักศึกษาที่ ตอบว่า “ใช่” สำหรับคำถามปกปิดเท่ากับ $130 - 75 = 55$ คน นั่นคือ จากนักศึกษาที่คาดว่าจะตอบ คำถามปกปิดทั้งหมด 150 คน มีนักศึกษาที่คาดว่าเคยทุจริตในการสอบจำนวน 55 คน และไม่เคย ทุจริตในการสอบจำนวน $150 - 55 = 95$ คน ดังรูปต่อไปนี้



4. ค่าประมาณสัดส่วนของนักศึกษาที่เคหะทุจริตในการสอบเท่ากับจำนวนนักศึกษาที่คาดว่าจะตอบว่า “ใช่” จำนวน 55 คน สำหรับคำถามปกปิด คือ “ท่านเคหะทุจริตในการสอบ ใช่หรือไม่” เทียบกับจำนวนนักศึกษาที่คาดว่าจะตอบคำถามปกปิด 150 คน ซึ่งเท่ากับ $\frac{55}{150} = 0.37$

3. สรุป

เทคนิคการเลือกตอบคำถามอย่างสุ่มเป็นวิธีการหนึ่งที่ใช้ในการสำรวจข้อมูลที่มีลักษณะปกปิดแทนวิธีการสัมภาษณ์โดยตรง เพื่อช่วยให้ผู้ถูกสัมภาษณ์สบายใจที่จะตอบคำถามตามความเป็นจริงมากขึ้น โดยที่ความลับไม่ถูกเปิดเผย และให้ค่าประมาณที่ใกล้เคียงกับค่าพารามิเตอร์มากขึ้น การใช้เทคนิคนี้จะบรรลุลวัตถุประสงค์ ถ้าผู้ศึกษากำหนดความน่าจะเป็นที่จะสุ่มได้คำถามปกปิดในเครื่องมือสุ่มเลือกในอัตราส่วนที่เหมาะสม ซึ่งผู้สนใจสามารถศึกษาได้จากงานวิจัยในเอกสารอ้างอิง

เอกสารอ้างอิง

- [1] Warner, S. L., 1965. Randomized Response : A Survey Technique for Eliminating Evasive Answer Bias. *Journal of the American Statistical Association*, 60, 63-69.
- [2] Greenberg, B.G., Abul-Ela, Abdel-Latif, A., Simmons, W.R., and Horvitz, D.G., 1969. The Unrelated Question Randomized Response Model: Theoretical Framework. *Journal of the American Statistical Association*, 60, 520-539.
- [3] Lanke, J., 1975. On the Choice of the Unrelated Question in Simmons Version of Randomized Response. *Journal of the American Statistical Association*, 70, 80-83.
- [4] Chaudhuri, A. and Mukerjee, R., 1988. Randomized Response: Theory and Techniques. Marcel Dekker, New York.
- [5] Mangat, N.S. and Singh R, 1990. An Alternative Randomized Response Procedure. *Biometrik*, 77, 439-442.
- [6] Kuk, A.Y.C., 1990. Asking Sensitive Questions Indirectly. *Biometrika*, 77, 436-438.
- [7] สรชัย พิศาลบุตร, 2524. การสำรวจข้อมูลที่มีการปกปิด. *เอกสารประกอบการประชุมวิชาการสถิติประยุกต์ ครั้งที่ 1*. กรุงเทพมหานคร : คณะสถิติประยุกต์ สถาบันบัณฑิตพัฒนบริหารศาสตร์, หน้า 1-23.

- [8] คลชาติ ตันติวานิช, 2532. การเปรียบเทียบวิธีการประมาณค่าสัดส่วนของประชากรที่มีลักษณะปกปิด. *วิทยานิพนธ์ปริญญาโทมหาบัณฑิต. ภาควิชาสถิติ บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.*
- [9] นุชนาถ คงช่วย, 2542. การศึกษาเทคนิคการตอบสนองอย่างสุ่มและการประยุกต์. *วิทยานิพนธ์ วิทยาศาสตร์มหาบัณฑิต. สาขาสถิติประยุกต์ มหาวิทยาลัยเชียงใหม่.*
- [10] Noodchanath Kongchay and Surin Khanabsakdi, 2002. Modification of Unrelated Question Randomized Response Model. *Chiang Mai J. Sci.*, 29(1), 49-50.
- [11] สุวรรณภา ปิ่นมณี และ สุรินทร์ ขนานศักดิ์, 2549. การพัฒนาวิธีของสรช้อยเกี่ยวกับเทคนิคการตอบสนองเชิงสุ่ม. *การประชุมวิชาการสถิติและสถิติประยุกต์ ประจำปี 2549, c75-c77.*
- [12] คลชาติ ตันติวานิช, 2550. การเปรียบเทียบเทคนิคการเลือกตอบคำถามอย่างสุ่ม 2 วิธี ซึ่งพัฒนาจากวิธีของสรช้อย. *เรื่องเต็มการประชุมวิชาการมหาวิทยาลัยเกษตรศาสตร์ ครั้งที่ 45 เล่มที่ 6 สาขาวิทยาศาสตร์* หน้า 406 -413.
- [13] คลชาติ ตันติวานิช, 2551. การขยายเทคนิคการเลือกตอบคำถามอย่างสุ่มซึ่งพัฒนาจากวิธีของสรช้อย. *เรื่องเต็มการประชุมวิชาการมหาวิทยาลัยเกษตรศาสตร์ ครั้งที่ 46 เล่มที่ 5 สาขาวิทยาศาสตร์,* หน้า 193 -200.
- [14] คลชาติ ตันติวานิช และ พรชัย หลายพสุ, 2552. การเปรียบเทียบเทคนิคการเลือกตอบคำถามอย่างสุ่ม 2 วิธี เสนอโดย Warner และ Kuk. *วารสารวิทยาศาสตร์ลาดกระบัง*, 18 (1), 13-21.