บทที่ 2

วิธีการประมาณค่าสัดส่วนของประชากรที่มีลักษณะปกปิดและผลงานวิจัยที่เกี่ยวข้อง

การประมาณค่าสัดส่วนของประชากรที่มีลักษณะปกปิดที่นำมาศึกษาครั้งที่มี 3 วิธีคือ

1. วิธีของวอร์เนอร์
2. วิธีของสรชัย
3. วิธีที่พัฒนาจากวิธีที่ 2

ซึ่งแต่ละวิธีแบ่งออกเป็น 2 กรณีคือกรณีที่ผู้ถูกสัมภาษณ์ทุกคนตอบตามความเป็นจริง (completely truthful) และกรณีที่บางส่วนของผู้ถูกสัมภาษณ์ที่มีลักษณะปกปิดตอบไม่ตรงตามความจริง (less than completely truthful)

2.1 วิธีของวอร์เนอร์

2.1.1 กรณีที่ผู้ถูกสัมภาษณ์ทุกคนตอบตามความจริง (completely truthful)

ในการประมาณค่าสัดส่วนของประชากรที่มีลักษณะปกปิดด้วยวิธีนี้ใช้ตัวอย่างชนิดแทนที่ขนาด n เครื่องมือสุ่มเลือก (randomizing device) ที่ให้ผู้ถูกสัมภาษณ์เลือกคำถามขึ้นมาตอบคือ ไพ่ 1 ชุด ซึ่งประกอบด้วยไพ่สีแดง และไพ่สีดำ ในอัตราส่วนที่ผู้สำรวจได้กำหนดไว้ล่วงหน้าแล้ว ในการตอบคำถามผู้สัมภาษณ์จะให้ผู้ถูกสัมภาษณ์สับไพ่จนเข้ากันดี เพื่อให้ไพ่แต่ละใบมีโอกาสถูกสุ่มด้วยความน่าจะเป็นเท่าๆ กัน แล้วดึงไพ่ 1 ใบ จากไพ่ชุดนั้นโดยไม่ให้ผู้สัมภาษณ์ทราบว่าเป็นไพ่สีอะไร ถ้าผู้ถูกสัมภาษณ์ดึงได้ไพ่สีแดง จะต้องตอบคำถามข้อแรกซึ่งเป็นเรื่องปกปิดที่ต้องการศึกษา เช่น "คุณเคยทำแท้ง ใช่หรือไม่" แต่ถ้าผู้ถูกสัมภาษณ์ดึงได้ไพ่สีดำ จะต้องตอบคำถามข้อสองซึ่งเป็นนิเสธของคำถามข้อแรก เช่น "คุณไม่เคยทำแท้ง ใช่หรือไม่" เมื่อผู้ถูกสัมภาษณ์ทุกคนเข้าใจในวิธีการนี้ และตอบคำถามตามความจริงแล้ว การประมาณค่าสัดส่วนของประชากรที่ลักษณะปกปิดด้วยวิธีภาวะน่าจะเป็นสูงสุด (maximum likelihood estimate) หาได้ดังนี้

ให้ π เป็นสัดส่วนของประชากรที่มีลักษณะปกปิด

r และ b เป็นจำนวนไพ่สีแดงและจำนวนไพ่สีดำในไพ่ 1 ชุด ตามลำดับ โดยที่ผู้สำรวจได้กำหนดไว้ล่วงหน้าแล้ว

p เป็นสัดส่วนของไพ่สีแดงในไพ่ 1 ชุดนั้น หรือ ความน่าจะเป็นที่จะสุ่มได้คำถามปกปิดที่ต้องการศึกษา

$$X_i = \begin{cases} 1 & \text{ถ้าผู้ถูกสัมภาษณ์คนที่ } i \text{ ตอบว่า "ใช่"} \\ 0 & \text{ถ้าผู้ถูกสัมภาษณ์คนที่ } i \text{ ตอบว่า "ไม่ใช่"} \end{cases}$$

ดังนั้น $pr\ [X_i = 1] = \pi p + (1-\pi)(1-p)$

$$pr\ [X_i = 0] = [1 - \{\pi p + (1-\pi)(1-p)\}]$$

ให้ n_1 แทนจำนวนผู้ถูกสัมภาษณ์ที่ตอบว่า "ใช่" จากจำนวนผู้ถูกสัมภาษณ์ทั้งหมด n คน จะได้ฟังก์ชันภาวะน่าจะเป็น (likelihood function)

$$L(\pi) = [\pi p + (1-\pi)(1-p)]^{n_1} [1 - \{\pi p + (1-\pi)(1-p)\}]^{n-n_1}$$

$$\log L(\pi) = n_1 \log [\pi p + (1-\pi)(1-p)] +$$

$$(n - n_1) \log [1 - \{\pi p + (1-\pi)(1-p)\}]$$

$$\frac{d}{d\pi} \log L(\pi) = \frac{n_1[p - (1-p)]}{\pi p + (1-\pi)(1-p)} + \frac{(n - n_1)[-p + (1-p)]}{[1 - \{\pi p + (1-\pi)(1-p)\}]}$$

$$= \frac{n_1(2p - 1)}{\pi p + (1-\pi)(1-p)} - \frac{(n - n_1)(2p - 1)}{[1 - \{\pi p + (1-\pi)(1-p)\}]}$$

$$\frac{n_1(2p - 1)}{\pi p + (1-\pi)(1-p)} - \frac{(n - n_1)(2p - 1)}{[1 - \{\pi p + (1-\pi)(1-p)\}]} = 0$$

$$\frac{n_1(2p - 1)}{\pi p + (1-\pi)(1-p)} = \frac{(n - n_1)(2p - 1)}{[1 - \{\pi p + (1-\pi)(1-p)\}]}$$

$$\frac{\left[1 - \{\pi p + (1-\pi)(1-p)\}\right]}{\pi p + (1-\pi)(1-p)} = \frac{n - n_1}{n_1}$$

$$= \frac{n}{n_1} - 1$$

$$\frac{\left[1 - \pi p + (1-\pi)(1-p)\right]}{\pi p + (1-\pi)(1-p)} + 1 = \frac{n}{n_1}$$

$$\frac{1}{\pi p + (1-\pi)(1-p)} = \frac{n}{n_1}$$

$$\pi p + (1-\pi)(1-p) = \frac{n_1}{n}$$

นั่นคือ ค่าประมาณภาวะน่าจะเป็นสูงสุด (maximum likelihood estimate) ของ π เท่ากับ

$$\hat{\pi}_w = \frac{p-1}{2p-1} + \frac{n_1}{n(2p-1)} \quad ; p \neq \frac{1}{2}$$

ค่าคาดหมาย (expected value) และความแปรปรวนของค่าประมาณ π คือ

$$E(\hat{\pi}_w) = \frac{p-1}{2p-1} + \frac{E(n_1)}{n(2p-1)}$$

$$= \frac{p-1}{2p-1} + \frac{n\left[\pi p + (1-\pi)(1-p)\right]}{n(2p-1)}$$

$$= \pi$$

$$\operatorname{Var}(\hat{\pi}_w) = \frac{\operatorname{Var}(n_1)}{n^2(2p-1)^2}$$

$$= \frac{[\pi p + (1-\pi)(1-p)]\,[1 - \{\pi p + (1-\pi)(1-p)\}]}{n(2p-1)^2}$$

$$= \frac{\pi(1-\pi)}{n} + \frac{p(1-p)}{n(2p-1)^2} \qquad ;\ p \neq \frac{1}{2}$$

นั่นคือ $\hat{\pi}_w$ เป็นค่าประมาณที่ไม่เอนเอียง (unbiased estimate) ของ π และความแปรปรวนของ $\hat{\pi}_w$ สามารถเขียนได้ในรูปของผลรวมระหว่างความแปรปรวนเนื่องจากการสุ่มตัวอย่าง (variance due to sampling) และความแปรปรวนเนื่องมาจากเครื่องมือสุ่มเลือก (variance due to randomizing device)

2.1.2 <u>กรณีที่บางส่วนของผู้ถูกสัมภาษณ์ที่มีลักษณะปกปิดตอบไม่ตรงตามความจริง</u> (less than completely truthful)

ให้ T_w แทนความน่าจะเป็นที่ผู้ถูกสัมภาษณ์ที่มีลักษณะปกปิดตอบคำถามปกปิดตามความจริง โดยที่ $0 < T_w \leq 1$

ดังนั้น $\operatorname{pr}[X_i=1] = \pi p T_w + \pi(1-p)(1-T_w) + (1-\pi)(1-p)$

$\operatorname{pr}[X_i=0] = [1 - \{\pi p T_w + \pi(1-p)(1-T_w) + (1-\pi)(1-p)\}]$

ให้ n_2 แทนจำนวนผู้ถูกสัมภาษณ์ที่ตอบว่า "ใช่" จากจำนวนผู้ถูกสัมภาษณ์ทั้งหมด n คน ค่าประมาณของ π เท่ากับ

$$\hat{\pi}'_w = \frac{p-1}{2p-1} + \frac{n_2}{n(2p-1)} \qquad ; \quad p \neq \frac{1}{2}$$

ค่าคาดหมาย (expected value) ของค่าประมาณ π คือ

$$E(\hat{\pi}'_w) = \frac{p-1}{2p-1} + \frac{E(n_2)}{n(2p-1)}$$

$$= \frac{p-1}{2p-1} + \frac{\pi pT_w + \pi(1-p)(1-T_w)+(1-\pi)(1-p)}{2p-1}$$

นั่นคือ $\hat{\pi}'_w$ เป็นค่าประมาณที่เอนเอียง (bias) ของ π
ดังนั้น ค่าความเอนเอียง และ ความแปรปรวนของค่าประมาณ π คือ

$$\text{Bias}\,(\hat{\pi}'_w) = E(\hat{\pi}'_w) - \pi$$

$$= \frac{p-1}{2p-1} + \frac{\pi pT_w + \pi(1-p)(1-T_w) + (1-\pi)(1-p)}{2p-1} - \pi$$

$$= \pi(T_w - 1)$$

$$\text{Var}(\hat{\pi}'_w) = \frac{\text{Var}(n_2)}{n^2(2p-1)^2}$$

$$= \frac{1}{n^2(2p-1)^2} \cdot \left[\pi pT_w + \pi(1-p)(1-T_w) + (1-\pi)(1-p)\right] \cdot \left[1 - \{\pi pT_w + \pi(1-p)(1-T_w)+(1-\pi)(1-p)\}\right]$$

$$= \frac{\pi T_w(1-\pi T_w)}{n} + \frac{p(1-p)}{n(2p-1)^2} \qquad ; \quad p \neq \frac{1}{2}$$

2.2 วิธีของสรชัย

2.2.1 กรณีที่ผู้ถูกสัมภาษณ์ทุกคนตอบตามความจริง (completely truthful)

การประมาณค่าสัดส่วนของประชากรที่มีลักษณะปกปิดด้วยวิธีการนี้ใช้ขนาดตัวอย่าง และเครื่องมือสุ่มเลือกเช่นเดียวกับวิธีของวอร์เนอร์ ในการตอบคำถาม ถ้าผู้ถูกสัมภาษณ์ดึงได้ไม่สีแดง จะต้องตอบคำถามข้อแรกซึ่งเป็นเรื่องปกปิดที่ต้องการศึกษา แต่ถ้าดึงได้ไม่สีดำ จะต้องตอบว่า "ใช่" เสมอ ซึ่งเหมือนกับถามคำถามที่สองว่า "ไม้ที่ท่านดึงได้เป็นไม้สีดำ ใช่หรือไม่" วิธีการนี้ก็คือรูปแบบหนึ่งของการใช้คำถามที่ไม่เกี่ยวข้องกันโดยกำหนดสัดส่วนของประชากรในคำถามที่สองให้เท่ากับ 1 เมื่อผู้ถูกสัมภาษณ์ทุกคนเข้าใจในวิธีการและตอบคำถามตามความจริงแล้ว การประมาณค่าสัดส่วนของประชากรที่มีลักษณะปกปิดด้วยวิธีภาวะน่าจะเป็นสูงสุด (maximum likelihood estimate) หาได้ดังนี้

$$Pr[X_i = 1] = \pi p + (1-p)$$

$$Pr[X_i = 0] = [1 - \{\pi p + (1-p)\}]$$

ให้ n_1 แทนจำนวนผู้ถูกสัมภาษณ์ที่ตอบว่า "ใช่" จากจำนวนผู้ถูกสัมภาษณ์ทั้งหมด n คน จะได้ฟังก์ชันภาวะน่าจะเป็น (likelihood function)

$$L(\pi) = [\pi p + (1-p)]^{n_1} [1 - \{\pi p + (1-p)\}]^{n-n_1}$$

$$\log L(\pi) = n_1 \log [\pi p + (1-p)] + (n-n_1) \log [1 - \{\pi p+(1-p)\}]$$

$$\frac{d}{d\pi}\log L(\pi) = \frac{n_1 p}{\pi p + (1-p)} - \frac{(n - n_1)\, p}{[1 - \{\pi p+(1-p)\}]}$$

$$-\frac{n_1 p}{\pi p + (1-p)} - \frac{(n - n_1)p}{[1 - \{\pi p + (1-p)\}]} = 0$$

$$\frac{n_1 p}{\pi p + (1-p)} = \frac{(n - n_1)p}{[1 - \{\pi p + (1-p)\}]}$$

$$\frac{[1 - \{\pi p + (1-p)\}]}{\pi p + (1-p)} = \frac{n - n_1}{n_1}$$

$$\frac{[1 - \{\pi p + (1-p)\}]}{\pi p + (1-p)} + 1 = \frac{n}{n_1}$$

$$\frac{1}{\pi p + (1-p)} = \frac{n}{n_1}$$

$$\pi p + (1-p) = \frac{n_1}{n}$$

นั่นคือ ค่าประมาณภาวะน่าจะเป็นสูงสุดของ π เท่ากับ

$$\hat{\pi}_s = 1 - \frac{1}{p} + \frac{n_1}{np}$$

ค่าคาดหมาย (expected value) และความแปรปรวนของค่าประมาณ π

คือ

$$E(\hat{\pi}_s) = 1 - \frac{1}{p} + \frac{E(n_1)}{np}$$

$$= 1 - \frac{1}{p} + \frac{n.\,[\pi p + (1-p)]}{np}$$

$$= \pi$$

$$Var(\hat{\pi}_s) = \frac{Var(n_1)}{n^2 p^2}$$

$$= \frac{n.\,[\pi p + (1-p)]\,[1 - \{\pi p + (1-p)\}]}{n^2 p^2}$$

$$= \frac{\pi(1-\pi)}{n} + \frac{(1-\pi)(1-p)}{np}$$

นั่นคือ $\hat{\pi}_s$ เป็นค่าประมาณที่ไม่เอนเอียง (unbiasid estimate) ของ π และความแปรปรวนของ $\hat{\pi}_s$ สามารถเขียนได้ในรูปของผลรวมระหว่างความแปรปรวนเนื่องจากการสุ่มตัวอย่างและความแปรปรวนเนื่องมาจากเครื่องมือสุ่มเลือก

2.2.2 กรณีที่บางส่วนของผู้ถูกสัมภาษณ์ที่มีลักษณะปกปิดตอบไม่ตรงตามความจริง (less than completely truthful)

ให้ T_s แทนความน่าจะเป็นที่ผู้ถูกสัมภาษณ์ที่มีลักษณะปกปิดตอบคำถามปกปิดตามความจริง โดยที่ $0 < T_s \leqslant 1$

ดังนั้น

$$\Pr\left[X_i = 1\right] = \pi p T_s + (1-p)$$

$$\Pr\left[X_i = 0\right] = \left[1 - \{\pi p T_s + (1-p)\}\right]$$

ให้ n_2 แทนจำนวนผู้ถูกสัมภาษณ์ที่ตอบว่า "ใช่" จากจำนวนผู้ถูกสัมภาษณ์ทั้งหมด n คน ค่าประมาณของ π เท่ากับ

$$\hat{\pi}_s' = 1 - \frac{1}{p} + \frac{n_2}{np}$$

ค่าคาดหมายของค่าประมาณ π คือ

$$E(\hat{\pi}_s') = 1 - \frac{1}{p} + \frac{E(n_2)}{np}$$

$$= 1 - \frac{1}{p} + \frac{n.\left[\pi p T_s + (1-p)\right]}{np}$$

$$= 1 - \frac{1}{p} + \frac{\left[\pi p T_s + (1-p)\right]}{p}$$

นั่นคือ $\hat{\pi}_s'$ เป็นค่าประมาณที่เอนเอียง (bias) ของ π

ดังนั้น ค่าความเอนเอียงและความแปรปรวนของค่าประมาณ π คือ

$$\text{Bias}(\hat{\pi}_s') = E(\hat{\pi}_s') - \pi$$

$$= 1 - \frac{1}{p} + \left|\frac{\pi p T_s + (1-p)}{p}\right| - \pi$$

$$= \pi(T_s - 1)$$

$$\text{Var}(\hat{\pi}_s') = \frac{\text{Var}(n_2)}{n^2p^2}$$

$$= \frac{n.\left[\pi pT_s+(1-p)\right]\left[1-\{\pi pT_s+(1-p)\}\right]}{n^2p^2}$$

$$= \frac{\pi T_s(1-\pi T_s)}{n} + \frac{(1-\pi T_s)(1-p)}{np}$$

2.3 <u>วิธีที่พัฒนาจากวิธีที่ 2</u>

2.3.1 <u>กรณีที่ผู้ถูกสัมภาษณ์ทุกคนตอบตามความจริง</u> (completely truthful)

การประมาณค่าสัดส่วนของประชากรที่มีลักษณะปกปิดด้วยวิธีนี้ใช้ขนาดตัวอย่างและเครื่องมือสุ่มเลือกเช่นเดียวกับวิธีของวอร์เนอร์ การตอบคำถามนั้น ผู้สัมภาษณ์จะให้ผู้ถูกสัมภาษณ์สับไพ่จนเข้ากันดี แล้วดึงไพ่ 2 ครั้ง ๆ ละใบแบบแทนที่ (with replacement) โดยไม่บอกให้ผู้สัมภาษณ์ทราบว่าดึงได้ไพ่สีอะไรในการดึงแต่ละครั้ง ถ้าผู้ถูกสัมภาษณ์ดึงได้ไพ่สีแดงในครั้งแรกและดึงได้ไพ่สีแดงหรือไพ่สีดำในครั้งที่สอง จะต้องตอบคำถามปกปิดที่ต้องการศึกษาตามความจริงว่า "ใช่" หรือ "ไม่ใช่" ถ้าผู้ถูกสัมภาษณ์ดึงได้ไพ่สีดำในครั้งแรกและไพ่สีแดงในครั้งที่สอง ผู้ถูกสัมภาษณ์จะต้องตอบว่า "ใช่" เสมอ แต่ถ้าดึงได้ไพ่สีดำในครั้งแรกและไพ่สีดำในครั้งที่สองจะต้องตอบว่า "ไม่ใช่" เสมอ ซึ่งเหมือนกับถามคำถามที่สองว่า ไพ่ที่ท่านดึงได้ใบแรกเป็นสีดำและใบที่สองเป็นสีแดง "ใช่หรือไม่" และสัดส่วนของประชากรในคำถามนี้เท่ากับความน่าจะเป็นที่สุ่มไพ่ใบแรกได้สีดำและสุ่มใบที่สองได้สีแดงต่อผลบวกของความน่าจะเป็นที่สุ่มไพ่ใบแรกได้สีดำและสุ่มใบที่สองได้สีแดงกับความน่าจะเป็นที่สุ่มใบแรกได้สีดำและใบที่สองได้สีดำ สำหรับค่าประมาณสัดส่วนของประชากรที่มีลักษณะปกปิดด้วยวิธีภาวะน่าจะเป็นสูงสุด หาได้ดังนี้

ให้ π เป็นสัดส่วนของประชากรที่มีลักษณะปกปิด

r และ b เป็นจำนวนไพ่สีแดงและไพ่สีดำในไพ่ 1 ชุด ตามลำดับ ซึ่งผู้สำรวจได้กำหนดไว้ล่วงหน้าแล้ว

p_r แทนความน่าจะเป็นที่สุ่มไพ่ใบแรกได้สีแดงและสุ่มใบที่สองได้ไพ่สีแดง หรือสีดำ เท่ากับความน่าจะเป็นที่สุ่มได้ไพ่สีแดง (p) ในวิธีของวอร์เนอร์หรือวิธีของสรชัยนั่นเอง

p_{br} แทนความน่าจะเป็นที่สุ่มไพ่ใบแรกได้สีดำและสุ่มใบที่สองได้ไพ่สีแดง

p_{bb} แทนความน่าจะเป็นที่สุ่มไพ่ใบแรกได้สีดำและสุ่มใบที่สองได้ไพ่สีดำ

ดังนั้น $Pr\ [X_i = 1] = \pi p + p_{br}$

$$Pr\ [X_i = 0] = [1 - \{\pi p + p_{br}\}]$$

ให้ n_1 แทนจำนวนผู้ถูกสัมภาษณ์ที่ตอบว่า "ใช่" จากจำนวนผู้ถูกสัมภาษณ์ทั้งหมด n คน จะได้ฟังก์ชันภาวะน่าจะเป็น (likelihood function)

$$L(\pi) = [\pi p + p_{br}]^{n_1} [1 - \{\pi p + p_{br}\}]^{n-n_1}$$

$$\log L(\pi) = n_1 \log [\pi p + p_{br}] + (n - n_1) \log [1 - \{\pi p + p_{br}\}]$$

$$\frac{d}{d\pi} \log L(\pi) = \frac{n_1 p}{\pi p + p_{br}} - \frac{(n - n_1) p}{[1 - \{\pi p + p_{br}\}]}$$

$$\frac{n_1 p}{\pi p + p_{br}} - \frac{(n - n_1) p}{[1 - \{\pi p + p_{br}\}]} = 0$$

$$\frac{n_1 p}{\pi p + p_{br}} = \frac{(n - n_1) p}{[1 - \{\pi p + p_{br}\}]}$$

$$\frac{[1 - \{\pi p + p_{br}\}]}{\pi p + p_{br}} = \frac{n - n_1}{n_1}$$

$$= \frac{n}{n_1} - 1$$

$$\frac{[1 - \{\pi p + p_{br}\}]}{\pi p + p_{br}} + 1 = \frac{n}{n_1}$$

$$\frac{1}{\pi p + p_{br}} = \frac{n}{n_1}$$

$$\pi p + p_{br} = \frac{n_1}{n}$$

นั่นคือ ค่าประมาณภาวะน่าจะเป็นสูงสุด (maximum likelihood estimate) ของ π เท่ากับ

$$\hat{\pi}_m = \frac{n_1}{np} - \frac{p_{br}}{p}$$

ค่าคาดหมายและความแปรปรวนของค่าประมาณ $\hat{\pi}$ คือ

$$E(\hat{\pi}_m) = \frac{E(n_1)}{np} - \frac{p_{br}}{p}$$

$$= \frac{n\,[\pi p + p_{br}]}{np} + \frac{p_{br}}{p}$$

$$= \pi$$

$$Var\ (\hat{\pi}_m) = \frac{Var(n_1)}{n^2p^2}$$

$$= \frac{n.\ [\pi p + (1-p)]\ [1 - \{\pi p + (1-p)\}]}{n^2p^2}$$

$$= \frac{\pi(1-p)}{n} + \frac{p_{br} - p_{br}^2 + \pi p - \pi p^2 - 2\pi p\, p_{br}}{np^2}$$

นั่นคือ $\hat{\pi}_m$ เป็นค่าประมาณที่ไม่เอนเอียง (unbiased estimate) ของ π และความแปรปรวนของ $\hat{\pi}_m$ สามารถเขียนได้ในรูปของผลรวมระหว่างความแปรปรวนเนื่องจากการสุ่มตัวอย่าง และความแปรปรวนเนื่องจากเครื่องมือสุ่มเลือก

2.3.2 <u>กรณีที่บางส่วนของผู้ถูกสัมภาษณ์ที่มีลักษณะปกปิดตอบไม่ตรงตามความจริง</u> (less than completely truthful)

ให้ T_m แทนความน่าจะเป็นที่ผู้ถูกสัมภาษณ์ที่มีลักษณะปกปิดตอบคำถามปกปิดตามความจริง โดยที่ $0 < T_m \leq 1$

ดังนั้น

$$Pr\ [X_i = 1] = \pi p T_m + p_{br}$$

$$Pr\ [X_i = 0] = [1 - \{\pi p T_m + p_{br}\}]$$

ให้ n_2 แทนจำนวนผู้ถูกสัมภาษณ์ที่ตอบว่า "ใช่" จากจำนวนผู้ถูกสัมภาษณ์ทั้งหมด n คน ค่าประมาณของ π เท่ากับ

$$\hat{\pi}'_m = \frac{n_2}{np} - \frac{p_{br}}{p}$$

ค่าคาดหมายของค่าประมาณ π คือ

$$E(\hat{\pi}'_m) = \frac{E(n_2)}{np} - \frac{p_{br}}{p}$$

$$= \frac{n\left[\pi pT_m + p_{br}\right]}{np} - \frac{p_{br}}{p}$$

นั่นคือ $\hat{\pi}'_m$ เป็นค่าประมาณที่เอนเอียงของ π

ดังนั้น ค่าความเอนเอียงและความแปรปรวนของค่าประมาณ π คือ

$$\text{Bias}(\hat{\pi}'_m) = E(\hat{\pi}'_m) - \pi$$

$$= \pi T_m - \pi$$

$$= \pi(T_m - 1)$$

$$\text{Var}(\hat{\pi}'_m) = \frac{\text{Var}(n_2)}{n^2p^2}$$

$$= \frac{n.\left[\pi pT_m + p_{br}\right]\left[1 - \{\pi pT_m + p_{br}\}\right]}{n^2p^2}$$

$$= \frac{\pi T_m(1-\pi T_m)}{n} + \frac{p_{br} - p_{br}^2 + \pi pT_m - \pi p^2T_m - 2\pi p\, p_{br}T_m}{np^2}$$

2.4. <u>ผลงานวิจัยที่เกี่ยวข้อง</u>

กรีนเบอร์กและคณะ (1969 : 520-539) ได้ศึกษาการประมาณค่าสัดส่วนของประชากรที่มีลักษณะปกปิดในเชิงทฤษฎี โดยใช้คำถามที่ไม่เกี่ยวข้องกันนั้น กรณีที่ทราบสัดส่วนของประชากรที่มีลักษณะทั่ว ๆ ไป ซึ่งไม่จำเป็นต้องปกปิด π_u จะใช้ตัวอย่างขนาด n เพียงตัวอย่างเดียวในการประมาณค่าสัดส่วนของประชากรที่มีลักษณะปกปิด (π) ค่าประมาณภาวะน่าจะเป็นสูงสุดและความแปรปรวนของค่าประมาณ π คือ

$$\hat{\pi} = \frac{\left[\hat{\phi} - (1-p)\,\pi_u\right]}{p}$$

$$\text{และ} \quad Var(\hat{\pi}) = \frac{\phi(1-\phi)}{np^2}$$

เมื่อ ϕ คือ ความน่าจะเป็นที่ผู้ถูกสัมภาษณ์จะตอบว่า "ใช่" ซึ่งเท่ากับ $p\pi + (1-p)\pi_u$

$\hat{\phi}$ คือ สัดส่วนของผู้ถูกสัมภาษณ์ที่ตอบว่า "ใช่" จากตัวอย่างทั้งหมด n คน

p คือ ความน่าจะเป็นที่จะสุ่มได้คำถามปกปิดที่ต้องการศึกษา

กรณีที่ไม่ทราบ π_u จะต้องใช้ตัวอย่าง 2 ชุด ขนาด n_1 และ n_2 ที่เป็นอิสระต่อกัน (independent) และไม่คาบเกี่ยวกัน (non-overlapping) โดยกำหนดให้ความน่าจะเป็นที่ผู้ถูกสัมภาษณ์ในตัวอย่างที่ 1 และตัวอย่างที่ 2 จะสุ่มได้คำถามปกปิดเป็น p_1 และ p_2 ตามลำดับ ค่าประมาณภาวะน่าจะเป็นสูงสุดและความแปรปรวนของค่าประมาณ π คือ

$$\hat{\pi} = \frac{\hat{\phi}_1(1-p_2) - \hat{\phi}_2(1-p_1)}{p_1 - p_2}$$

$$Var(\hat{\pi}) = \frac{1}{(p_1-p_2)^2}\left[\frac{\phi_1(1-\phi_1)(1-p_2)^2}{n_1} + \frac{\phi_2(1-\phi_2)(1-p_2)^2}{n_2}\right]$$

เมื่อ ϕ_1 และ ϕ_2 ความน่าจะเป็นที่ผู้ถูกสัมภาษณ์ในตัวอย่างที่ 1 และตัวอย่างที่ 2 จะตอบว่า "ใช่" ซึ่งเท่ากับ

$$\phi_1 = p_1\pi + (1-p_1)\,\pi_u$$

$$\phi_2 = p_2\pi + (1-p_2)\,\pi_u$$

$\hat{\phi}_1$ และ $\hat{\phi}_2$ คือสัดส่วนของผู้ถูกสัมภาษณ์ที่ตอบว่า "ใช่
ในตัวอย่างที่ 1 และ ตัวอย่างที่ 2 ตามลำดับ

การศึกษาครั้งนี้ เขาพบว่าค่าประมาณที่ได้อาจตกอยู่นอกช่วง [0 ,1] ถ้า $p_1 - p_2$ แตกต่างกันน้อย และ n มีขนาด เล็ก ดังนั้น จึงควรเลือก $p_1 - p_2$ ให้แตกต่างกันมากที่สุด โดยให้ p_1 เข้าใกล้ 1 มากที่จะไม่ทำให้ผู้ถูกสัมภาษณ์เกิดความไม่ไว้วางใจแล้วจึงเลือก p_2 ให้ห่างจาก p_1 มากที่สุด ในทางปฏิบัติเขาแนะนำให้เลือก p_1 อยู่ในช่วง 0.8 $\pm$ 0.1 แล้วเลือก p_2 โดยมีข้อจำกัดว่า $p_1 + p_2 = 1$ สำหรับความแปรปรวนของตัวประมาณนั้น เขาพบว่า ถ้า $p_1 > 0.5$ แล้ว var $(\hat{\pi})$ จะมีค่าต่ำสุด เมื่อ $p_2 = 0$ ซึ่งตรงกับผลการศึกษาของมัวร์ (1971 : 627-629) แต่ในทางปฏิบัติเขาแนะนำให้ใช้กฎ $p_1 + p_2 = 1$ เพราะ การใช้ $p_2 = 0$ อาจทำให้ผู้ถูกสัมภาษณ์ให้ความร่วมมือน้อยลง

ดาวลิ่งและชัชท์แมน (1975 : 84-89) แสดงให้เห็นว่าความแปรปรวนของตัวประมาณที่ใช้คำถามซึ่งไม่เกี่ยวข้องกันมีค่าต่ำกว่าความแปรปรวนของตัวประมาณที่ใช้คำถามซึ่งเกี่ยวข้องกัน เมื่อ p หรือ max (p_1 , p_2) ในกรณีใช้ 2 ตัวอย่างมีค่ามากกว่า $\frac{1}{3}$ โดยประมาณ สำหรับทุกค่า π และ π_u