

บทที่ 2

แนวคิด ทฤษฎี และผลงานวิจัยที่เกี่ยวข้อง

2.1 ความหมายของการประมาณค่าและตัวประมาณ [2]

ในการศึกษาลักษณะบางอย่างของประชากร สมมติให้ลักษณะนั้นแทนด้วยตัวแปรสุ่ม X ที่มีฟังก์ชันความน่าจะเป็น (probability density function, p.d.f.) เป็น $f(x; \pi)$ ที่ขึ้นอยู่กับพารามิเตอร์ π ซึ่งเป็นค่าคงตัวที่ไม่ทราบค่า และเราต้องการหาค่าพารามิเตอร์นี้ ในทางปฏิบัติทำได้โดยสุ่มตัวอย่างมาจำนวนหนึ่ง สมมติว่าตัวอย่างมีขนาด n ให้ เป็น $x_1, x_2, x_3, \dots, x_n$ เป็นค่าสังเกตที่ได้จากตัวอย่างเชิงสุ่ม $X_1, X_2, X_3, \dots, X_n$ จากประชากรที่มี p.d.f. เป็น $f(x; \pi)$ และใช้ค่าสังเกตเหล่านั้นมาทำการประมาณค่า π ซึ่งในการประมาณค่าพารามิเตอร์ทำได้สองแบบ แบบแรกเรียกว่า การประมาณค่าแบบจุด (Point Estimation) หรือการประมาณค่าด้วยค่าคงตัว กล่าวคือถ้าใช้ค่าของสถิติบางตัวเช่น $Y_1 = u(X_1, X_2, X_3, \dots, X_n)$ เป็นค่าประมาณของ π จะเรียกสถิติ $Y_1 = u(X_1, X_2, X_3, \dots, X_n)$ ว่าตัวประมาณแบบจุด เช่น การประมาณค่าระดับสติปัญญาเฉลี่ยของเด็กนักเรียนระดับมัธยมมีค่าเท่ากับ 100 โดยประมาณเฉลี่ยจากตัวอย่างนักเรียนมัธยม 100 คน เป็นต้นซึ่งตัวเลขนี้จะเป็นค่าคงตัวที่นำไปประมาณค่าพารามิเตอร์ ส่วนการประมาณอีกแบบเรียกว่า การประมาณแบบช่วง (Interval Estimation) กล่าวคือถ้ากำหนด $Y_1 = u_1(X_1, X_2, X_3, \dots, X_n)$ และ $Y_2 = u_2(X_1, X_2, X_3, \dots, X_n)$ เป็นค่าประมาณของ π โดยที่ $u_1(X_1, X_2, X_3, \dots, X_n) < u_2(X_1, X_2, X_3, \dots, X_n)$ และ $(u_1(X_1, X_2, X_3, \dots, X_n), u_2(X_1, X_2, X_3, \dots, X_n))$ จะเป็นช่วงที่มี π อยู่ เรียก $(u_1(X_1, X_2, X_3, \dots, X_n), u_2(X_1, X_2, X_3, \dots, X_n))$ ว่าตัวประมาณแบบช่วงของ π เช่นในกรณีเดียวกันระดับสติปัญญาเฉลี่ยของเด็กนักเรียนระดับมัธยมมีค่าเท่ากับ 100 เราอาจประมาณเป็นว่าระดับสติปัญญาของนักเรียนมัธยมจะเป็น 80-120 ด้วยช่วงความเชื่อมั่น 95 %ก็ได้ ในการศึกษาครั้งนี้จะกล่าวถึงเฉพาะการประมาณแบบจุดเท่านั้น

นิยาม 2.1 การประมาณแบบจุดคือ วิธีการเลือกสถิติที่เหมาะสมเพื่อใช้ค่าของสถิตินั้นทำการประมาณค่าพารามิเตอร์ π เรียกสถิติใดๆซึ่งค่าของมันได้รับการนำไปใช้ในการประมาณค่าพารามิเตอร์ π ว่าตัวประมาณ (Estimator) ของ π

2.2 หลักเกณฑ์ในการเลือกตัวประมาณ [9]

ในการเลือกตัวประมาณล้วนแต่ต้องการตัวประมาณที่ดี ในแง่ความถูกต้องใกล้เคียงกับค่าจริงๆ ของประชากรทั้งสิ้น หลักการที่จะตัดสินใจว่าตัวประมาณใดเป็นตัวประมาณที่ดีและวิธีการหาค่าตัวประมาณต่างๆ จึงเป็นเรื่องสำคัญอย่างยิ่งในทางสถิติ ทั้งนี้ในการเลือกตัวประมาณนั้น ย่อมต้องขึ้นอยู่กับจุดประสงค์ในแต่ละสถานการณ์นั้นๆ ด้วย จะไม่มีหลักเกณฑ์ที่ใช้ได้เพียงเกณฑ์เดียวเท่านั้น ส่วนใหญ่จะอาศัยคุณสมบัติของสถิติหลายเกณฑ์ในการที่จะถือว่า ตัวสถิตินั้นเป็นตัวสถิติที่ดี โดยมีหลักเกณฑ์ต่างๆ ไปดังนี้

2.2.1 ตัวประมาณที่มีความไม่เอนเอียง (Unbiased Estimator)

2.2.2 ตัวประมาณที่มีความคงเส้นคงวา (Consistent Estimator)

2.2.3 ตัวประมาณที่มีความแปรปรวนต่ำสุด (Minimum Variance Estimator)

2.2.4 ตัวประมาณที่มีประสิทธิภาพ (Efficient Estimator)

2.2.5 ตัวประมาณที่มีความพอเพียง (Sufficient Estimator)

2.2.1 ตัวประมาณที่มีความไม่เอนเอียง

นิยาม 2.2 ให้ $\hat{\pi} = u(x_1, x_2, \dots, x_n)$ เป็นตัวประมาณของพารามิเตอร์ π และ $\hat{\pi}$ เป็นตัวประมาณที่ไม่เอนเอียงของ π ก็ต่อเมื่อ $E(\hat{\pi}) = \pi$ นั่นคือถ้า $\hat{\pi}$ เป็นตัวประมาณที่ไม่เอนเอียงต่อ π จะได้ว่าค่าคาดหวังของ $\hat{\pi}$ คือค่าเดียวกันกับค่า π นั่นเอง

ตัวอย่าง 1 จงแสดงว่า $\hat{\pi} = \sum_{i=1}^n a_i x_i$ (โดยที่ $\sum_{i=1}^n a_i = 1$) เป็นตัวประมาณที่ไม่เอนเอียงของค่าเฉลี่ย π ของประชากรใดๆ

วิธีทำ กำหนดให้ $x_1, x_2, \dots, x_n$ เป็นตัวอย่างสุ่มขนาด n จากประชากรที่มีค่าเฉลี่ย π

$$\begin{aligned} E(X) &= \pi \\ E(\hat{\pi}) &= E \sum_{i=1}^n a_i x_i \\ &= \sum_{i=1}^n a_i E(X) \\ &= \sum_{i=1}^n a_i \pi \\ &= \pi \sum_{i=1}^n a_i \\ &= \pi \end{aligned}$$

นิยาม 2.3 ถ้า $\hat{\pi}$ เป็นตัวประมาณของ π ค่าความเอนเอียง (Bias) ของตัวประมาณ คือ $\text{Bias} = E(\hat{\pi}) - \pi$

นิยาม 2.4 ความคลื่อนคลาดของการประมาณค่า คือระยะระหว่างทางระหว่างตัวประมาณกับพารามิเตอร์นั้นคือ ความคลื่อนคลาด $= |\hat{\pi} - \pi|$ [2]

โดยทั่วไปมักใช้ค่าเฉลี่ยของกำลังสองของความคลื่อนคลาด (Mean Square Error) แทนด้วย MSE เป็นตัววัด $MSE(\hat{\pi}) = E[(\hat{\pi} - \pi)^2]$

นิยาม 2.5 ให้ $\hat{\pi}$ เป็นตัวประมาณแบบจุดของ π เราเรียก $E[(\hat{\pi} - \pi)^2]$ ว่าค่าเฉลี่ยความคลื่อนคลาด กำลังสองของตัวประมาณ π

$$\text{หมายเหตุ} \quad MSE(\hat{\pi}) = \text{Var}(\hat{\pi}) + [B(\hat{\pi})]^2$$

ซึ่งพิสูจน์ได้ดังนี้

$$MSE(\hat{\pi}) = E[(\hat{\pi} - \pi)^2]$$

$$= E[(\hat{\pi} - E(\hat{\pi}) + E(\hat{\pi}) - \pi)^2]$$

$$= E[\hat{\pi} - E(\hat{\pi})]^2 + [E(\hat{\pi}) - \pi]^2$$

$$= \text{Var}(\hat{\pi}) + [B(\hat{\pi})]^2$$

ถ้า $\hat{\pi}$ เป็นตัวประมาณที่ไม่เอนเอียงของ π แล้ว $\text{Bias}(\hat{\pi}) = 0$ จะได้ว่า $MSE(\hat{\pi}) = \text{Var}(\hat{\pi})$

2.2.2 ตัวประมาณที่มีความคงเส้นคงวา

สมมติโยนเหรียญอันหนึ่ง n ครั้ง ในการโยนแต่ละครั้งจะเป็นอิสระแก่กันให้ X เป็นจำนวนครั้งที่เหรียญขึ้นหัวในการโยน n ครั้ง ซึ่ง X จะมีการแจกแจงแบบทวินาม โดยมี p เป็นความน่าจะเป็นที่เหรียญขึ้นหัวเป็นพารามิเตอร์ถ้าไม่ทราบค่าที่แท้จริงของ p สัดส่วนของตัวอย่างคือ

$\frac{X}{n}$ เป็นตัวประมาณของ p ซึ่งถ้าจำนวนครั้งในการโยนเหรียญมากขึ้น $\frac{X}{n}$ ควรจะเข้าใกล้ p ที่แท้จริง

จริง นั่นคือเมื่อขนาดตัวอย่างมากขึ้นตัวประมาณควรจะเข้าใกล้พารามิเตอร์ที่ต้องการประมาณค่า

$\left| \frac{X}{n} - p \right|$ จะน้อยกว่าจำนวนจริงบวกซึ่งเป็นสมาชิกใดๆ ถ้า $\frac{X}{n}$ เข้าใกล้ p เมื่อ n มีค่ามากๆ

$\Pr \left[\left| \frac{X}{n} - p \right| \leq \epsilon \right]$ ควรมีค่าเข้าใกล้ 1 และเมื่อ n มีค่ามากขึ้น หรือ $\lim_{n \rightarrow \infty} \Pr \left[\left| \frac{X}{n} - p \right| \leq \epsilon \right] = 1$ ใน

กรณีนี้กล่าวว่า $\frac{X}{n}$ เป็นตัวประมาณที่คงเส้นคงวา

นิยาม 2.6 ตัวประมาณ $\hat{\pi}_n$ ของ π ที่คิดมาจากตัวอย่างสุ่มขนาด n เป็นตัวประมาณที่คงเส้นคงวาของ π ถ้า $\hat{\pi}_n$ ลู่เข้าสู่ความน่าจะเป็นไปสู่ π นั่นคือเมื่อ ε เป็นจำนวนจริงบวก $\lim_{n \rightarrow \infty} \Pr[|\hat{\pi}_n - \pi| > \varepsilon] = 0$ หรือ $\lim_{n \rightarrow \infty} \Pr[|\hat{\pi}_n - \pi| \leq \varepsilon] = 1$

ทฤษฎี 2.1 ถ้าตัวประมาณ $\hat{\pi}_n$ ของ π โดยใช้ตัวอย่างสุ่มขนาด n ถ้า $\lim_{n \rightarrow \infty} E(\hat{\pi}_n) = \pi$ และ $\lim_{n \rightarrow \infty} \text{Var}(\hat{\pi}_n) = 0$ แล้ว $\hat{\pi}_n$ จะเป็นตัวประมาณที่คงเส้นคงวาของ π

ตัวอย่าง 2 ให้ X เป็นตัวแปรแทนค่าสังเกตใด ๆ ที่แจกแจงแบบปกติด้วยค่าเฉลี่ยและความแปรปรวน เท่ากับ μ และ σ^2 ตามลำดับสุ่มตัวอย่างขนาด n จากประชากรดังกล่าวให้เป็น $x_1, x_2, x_3, \dots, x_n$ จงแสดงว่า ค่าเฉลี่ยของ $\bar{X}$ เป็นตัวประมาณที่คงเส้นคงวาของพารามิเตอร์ μ

$$\begin{aligned} \text{วิธีทำ } \bar{X} &\sim N\left(\mu, \frac{\sigma^2}{n}\right) \\ \lim_{n \rightarrow \infty} E(\bar{X}) &= \mu \\ \lim_{n \rightarrow \infty} \text{Var}(\bar{X}) &= \lim_{n \rightarrow \infty} \left(\frac{\sigma^2}{n}\right) \\ &= 0 \end{aligned}$$

ดังนั้น $\bar{X}$ เป็นตัวประมาณที่คงเส้นคงวาต่อพารามิเตอร์ μ

2.2.3 ตัวประมาณที่ไม่เอนเอียงและมีความแปรปรวนต่ำที่สุด

นิยาม 2.9 ให้ $x_1, x_2, \dots, x_n$ เป็นตัวอย่างสุ่มจากประชากรที่มีฟังก์ชันความหนาแน่น $f(x; \pi)$ ถ้าให้ $\hat{\pi}$ เป็นตัวประมาณสำหรับพารามิเตอร์ π โดยที่

1. $E(\hat{\pi}) = \pi$ นั่นคือ $\hat{\pi}$ เป็นตัวประมาณที่ไม่เอนเอียงของ π
2. $\text{Var}(\hat{\pi})$ น้อยกว่าความแปรปรวนของตัวประมาณไม่เอนเอียงอื่น ๆ ของ π

แล้ว $\hat{\pi}$ เป็นตัวประมาณที่ไม่เอนเอียงและความแปรปรวนต่ำที่สุดของ π หรืออาจกล่าวได้ว่า $\hat{\pi}$ เป็นตัวสถิติที่ดีที่สุด

2.2.4 ตัวประมาณที่มีประสิทธิภาพ

ในกรณีที่มีตัวประมาณที่ไม่เอนเอียงมากกว่าหนึ่งตัวเป็นตัวประมาณของพารามิเตอร์ π สมมติ $\hat{\pi}_1$ และ $\hat{\pi}_2$ เป็นตัวประมาณที่ไม่เอนเอียงของ π จะเลือกใช้ตัวประมาณที่มีความแปรปรวนน้อยกว่า นั่นคือ ถ้า $\hat{\pi}_1$ และ $\hat{\pi}_2$ เป็นตัวประมาณที่ไม่เอนเอียงของ π แล้ว $\hat{\pi}_1$ จะมีประสิทธิภาพมากกว่า $\hat{\pi}_2$ ก็ต่อเมื่อ $\text{Var}(\hat{\pi}_1) < \text{Var}(\hat{\pi}_2)$

นิยาม 2.7 $\hat{\pi}_1$ และ $\hat{\pi}_2$ เป็นตัวประมาณที่ไม่เอนเอียงของ π โดยมีความแปรปรวนเป็น $\text{Var}(\hat{\pi}_1)$ และ $\text{Var}(\hat{\pi}_2)$ ตามลำดับแล้วประสิทธิภาพสัมพัทธ์ (Relative Efficiency) ของตัวประมาณ $\hat{\pi}_1$ เทียบกับตัวประมาณ $\hat{\pi}_2$ ซึ่งแทนด้วย $e(\hat{\pi}_1, \hat{\pi}_2)$ คือ

$$e(\hat{\pi}_1, \hat{\pi}_2) = \frac{\text{Var}(\hat{\pi}_2)}{\text{Var}(\hat{\pi}_1)}$$

ถ้าประสิทธิภาพสัมพัทธ์ของ $\hat{\pi}_1$ เทียบ $\hat{\pi}_2$ มากกว่า 1 แสดงว่า $\hat{\pi}_1$ เป็นตัวประมาณที่ดีกว่า $\hat{\pi}_2$ โดยทั่วไปในการเปรียบเทียบจะมีความหมายยิ่งขึ้น มักจะเปรียบเทียบประสิทธิภาพกับตัวประมาณที่มีความแปรปรวนต่ำที่สุดในบรรดาตัวประมาณที่ไม่เอนเอียงด้วยกัน เช่นถ้า $\hat{\pi}_1$ เป็นตัวประมาณที่ไม่เอนเอียงและมีความแปรปรวนต่ำที่สุด และ $\hat{\pi}_2$ เป็นตัวประมาณที่ไม่เอนเอียงที่สนใจ ถ้า

$$e(\hat{\pi}_1, \hat{\pi}_2) = \frac{\text{Var}(\hat{\pi}_1)}{\text{Var}(\hat{\pi}_2)} = 1 \text{ กล่าวได้ว่า } \hat{\pi}_2 \text{ เป็นตัวประมาณที่มีประสิทธิภาพสูงสุดเพราะ } \hat{\pi}_2 \text{ มี}$$

ความแปรปรวนต่ำที่สุดเท่ากับความแปรปรวนของ $\hat{\pi}_1$ นั่นเอง

นิยาม 2.8 ตัวประมาณ $\hat{\pi}$ สำหรับพารามิเตอร์ π เป็นตัวประมาณที่มีประสิทธิภาพถ้าตัวประมาณ $\hat{\pi}$ มีความแปรปรวนน้อยที่สุดในบรรดาตัวประมาณที่ไม่เอนเอียงของ π

2.2.5 ตัวประมาณที่มีความพอเพียง

นิยาม 2.10 ให้ $X_1, X_2, X_3, \dots, X_n$ เป็นตัวอย่างเชิงสุ่มขนาด n จากประชากรที่มี p.d.f. เป็น $f(x; \pi), \pi \in \Omega$ ให้ $Y_1 = u(X_1, X_2, X_3, \dots, X_n)$ เป็นตัวสถิติที่มี p.d.f. เป็น $g_1(y_1; \pi)$ แล้ว Y_1 จะเป็นตัวสถิติที่เพียงพอสำหรับ π ก็ต่อเมื่อ

$$\frac{f(x_1; \pi)f(x_2; \pi)f(x_3; \pi) \dots f(x_n; \pi)}{g_1[u_1(x_1, x_2, \dots, x_n)]} = H(x_1, x_2, \dots, x_n)$$

เท่านั้นเมื่อ $H(x_1, x_2, \dots, x_n)$ เป็นฟังก์ชันที่ไม่ขึ้นกับ π สำหรับทุกๆค่าที่ $y_1 = u(x_1, x_2, \dots, x_n)$ ที่กำหนดให้

ตัวอย่าง 3 ให้ $X_1, X_2, X_3, \dots, X_n$ เป็นตัวอย่างสุ่มขนาด n จากประชากรที่มี p.d.f. คือ

$$\begin{aligned} f(x; \pi) &= \pi^x (1 - \pi)^{1-x}, x = 0, 1, 0 < \pi < 1 \\ &= 0 \quad \text{เมื่อ } x \text{ เป็นค่าอื่นๆ} \end{aligned}$$

จงแสดงว่าสถิติ $Y_1 = X_1 + X_2 + X_3 + \dots + X_n$ เป็นสถิติที่พอเพียงสำหรับ π

วิธีทำ
$$Y_i = \sum_{i=1}^n x_i$$

เนื่องจาก $y_i = x_1 + x_2 + x_3 + \dots + x_n$ ซึ่งเท่ากับจำนวนครั้งของ 1's ในการกระทำ n ครั้งที่อิสระต่อกัน ดังนั้น Y_i มีการแจกแจงแบบทวินาม โดยมีพารามิเตอร์ n และ π และ p.d.f. ของ Y_i คือ

$$g_1(y_i; \pi) = {}^n C_{y_i} (\pi)^{y_i} (1-\pi)^{n-y_i}, \quad y_i = 0, 1, 2, \dots, n$$

$$= 0 \quad \text{เมื่อ } y_i \text{ เป็นค่าอื่น}$$

$$\frac{f(x_1; \pi) \dots f(x_n; \pi)}{g_1(y_i; \pi)} = \frac{\pi^{x_1} (1-\pi)^{1-x_1} \pi^{x_2} (1-\pi)^{1-x_2} \dots \pi^{x_n} (1-\pi)^{1-x_n}}{{}^n C_{y_i} \pi^{y_i} (1-\pi)^{n-y_i}}$$

$$= \frac{\pi^{\sum x_i} (1-\pi)^{n-\sum x_i}}{{}^n C_{\sum x_i} \pi^{\sum x_i} (1-\pi)^{n-\sum x_i}}$$

$$= \frac{1}{{}^n C_{\sum x_i}}$$

ซึ่งไม่ขึ้นอยู่กับ π ดังนั้น Y_i เป็นสถิติที่เพียงพอสำหรับ π

2.3 วิธีการหาตัวประมาณแบบภาวะน่าจะเป็นสูงสุด [9]

วิธีการประมาณค่าวิธีนี้มาจากความคิดที่ว่าประชากรที่แตกต่างกัน ทำให้เกิดตัวอย่างที่แตกต่างกัน และตัวอย่างที่กำหนดให้จะมาจากประชากรบางจำนวน พิจารณากรณีที่มีประชากรมีการกระจายแบบปกติและตัวอย่างที่กำหนดให้มีขนาด n ตัวอย่างจะเป็นจุดซึ่งกระจายอยู่บนแกนรอบ ๆ ค่าเฉลี่ย โดยทั่ว ๆ ไปแล้วความแตกต่างของประชากรที่มีการกระจายแบบปกติเนื่องมาจากค่าเฉลี่ยและความแปรปรวนที่ต่างกัน ถ้าพิจารณาประชากรที่มีความแปรปรวนเท่ากัน กลุ่มตัวอย่างที่มีค่าเฉลี่ย 5 จะมาจากประชากรที่มีค่าเฉลี่ย 6 หรือ 7 ดังนั้น ตัวอย่างที่มีความแปรปรวนใหญ่จะมาจากประชากรที่มีความแปรปรวนใหญ่มากกว่าที่จะมาจากประชากรที่มีค่าความแปรปรวนเล็ก ดังนั้นจึงต้องพิจารณาค่าเฉลี่ยและความแปรปรวนของประชากรที่เกี่ยวข้องกับค่าเฉลี่ยและความแปรปรวนของตัวอย่าง

นิยาม 2.11 ฟังก์ชันภาวะน่าจะเป็น (Likelihood function) ของตัวอย่างสุ่มได้แก่ ฟังก์ชันภาวะน่าจะเป็นร่วมของ $x_1, x_2, \dots, x_n$ โดยถือว่าเป็นฟังก์ชันของภาวะน่าจะเป็นของพารามิเตอร์ π โดยแทนฟังก์ชันภาวะน่าจะเป็น ด้วย L หรือ $L(x_1, x_2, \dots, x_n; \pi)$ หรือ $L(\pi)$ นั่นคือ

$$L(x_1, x_2, \dots, x_n; \pi) = f(x_1; \pi), \dots, f(x_n; \pi)$$

$$= \prod_{i=1}^n f(x_i; \pi) \quad \text{โดยถือว่า } L \text{ เป็นฟังก์ชันของ } \pi$$

นิยาม 2.12 ตัวประมาณแบบภาวะน่าจะเป็นสูงสุด (Maximum Likelihood Estimator) แทนด้วย MLE ของ π ได้แก่ $\hat{\pi} = \hat{\pi}(x_1, x_2, \dots, x_n)$ ที่เป็นค่าของ π ในสเปซของพารามิเตอร์ ที่ทำให้ฟังก์ชันภาวะน่าจะเป็นมีค่าสูงสุด

วิธีการหาตัวประมาณแบบภาวะน่าจะเป็นสูงสุด มีข้อสังเกตดังนี้

- 1) จะใช้ค่าของ π ที่ทำให้ $L(x_1, x_2, \dots, x_n; \pi)$ มีค่ามากที่สุด. ให้เป็น π นั่นคือ

$$L(x_1, x_2, \dots, x_n; \pi) \geq L(x_1, x_2, \dots, x_n; \pi), \pi \in \Omega$$

- 2) ในหลายกรณีเราอาจจะใช้อนุพันธ์ในการหา π ที่ทำให้ $L(\pi) \geq L(\pi)$ เมื่อ $\pi \in \Omega$ ถ้าเรารู้ฟังก์ชัน $f(x; \pi)$ ไม่ขึ้นอยู่กับการหา π และ π อาจจะมีค่าในช่วงช่วงหนึ่งค่าวิกฤติของฟังก์ชันภาวะน่าจะเป็น (ถ้ามี) ได้แก่รากของสมการ $\frac{\partial L}{\partial \pi} = 0$ และเงื่อนไขที่พอเพียงที่ π จะให้ค่า

สูงสุดของ $L(x_1, x_2, \dots, x_n; \pi)$ ได้แก่ $\frac{\partial^2 L}{\partial \pi^2} < 0$ เมื่อ $\pi = \hat{\pi}$

- 3) การใช้อนุพันธ์ หาตัวประมาณแบบภาวะน่าจะเป็นสูงสุดในหลายกรณีหากใช้ $\ln L$ แทน L จะสะดวกกว่าเราเรียกสมการ $\frac{\partial L}{\partial \pi} = 0$ หรือ $\frac{\partial \ln L}{\partial \pi} = 0$ ว่าสมการภาวะน่าจะเป็น (Likelihood equation) ซึ่งเป็นสมการที่ใช้ในการหาตัวประมาณแบบภาวะน่าจะเป็นสูงสุด $\hat{\pi}$ เรียก $\hat{\pi}$ ว่าตัวสถิติภาวะน่าจะเป็นสูงสุด (Maximum likelihood estimator)

- 4) การหาตัวประมาณแบบภาวะน่าจะเป็นสูงสุด ในบางกรณีอาจจะใช้อนุพันธ์ไม่ได้ซึ่งจำเป็นต้องพิจารณาด้วยเทคนิคด้วยเทคนิคทางคณิตศาสตร์อย่างอื่น

นอกจากนี้ยังมีอีกหลายวิธีในการหาตัวประมาณแต่ไม่ขอกล่าวถึงรายละเอียดในที่นี้เช่น

วิธีโมเมนต์ (Method of moment)

วิธีไคกำลังสองต่ำสุด (Method of minimum chi-square)

วิธีของเบย์ (Method of Bayes)

วิธีมินิแมกซ์ (Method of minimax)

วิธีการของพิทมัน (Method of Pitman) เป็นต้น

2.4 เทคนิคการตอบสนองการสุ่มสำหรับข้อมูลเชิงคุณภาพ

เทคนิคการตอบสนองการสุ่ม (Randomized Response Technique) เป็นเทคนิคที่ใช้ในการเก็บข้อมูลเมื่อข้อมูลที่สนใจเกี่ยวข้องกับพฤติกรรมของมนุษย์ ในเรื่องที่ทำให้ผู้ตอบรู้สึกถูกรบกวน และสร้างความลำบากใจมีผลทำให้ผู้ตอบปฏิเสธหรือหลีกเลี่ยงที่จะให้ความร่วมมือโดยไม่ตอบคำถามหรือไม่ให้ข้อมูลที่แท้จริง เทคนิคการตอบสนองการสุ่มนี้จึงออกแบบมาเพื่อ

ทำให้ผู้ตอบให้ความร่วมมือต่อการตอบคำถามและให้ตอบความจริงอย่างสบายใจ จึงได้มีนักสถิติหลายท่าน ได้สร้างเครื่องมือและพัฒนาเครื่องมือสำหรับการรวบรวมข้อมูลที่มีลักษณะดังกล่าวต่อไปนี้

2.4.1 วอร์เนอร์

2.4.2 กรีนเบิร์กและคณะ

2.4.3 สรชัย

2.4.4 ดลชาติ

2.4.5 โพลสันและคณะ

2.4.1 วอร์เนอร์ [31]

วอร์เนอร์เป็นบุคคลแรกที่คิดค้นเทคนิคนี้ แนวคิดของเขาคือทำอย่างไรให้ผู้ตอบรู้สึกสบายใจในการที่จะตอบความจริงคำถามปกปิด ที่สนใจด้วยความจริง ในที่นี้แทนคำถามดังกล่าวด้วย Q_A เช่น Q_A : “ท่านเคยเสพยาบ้า ใช่หรือไม่” ซึ่งคำถามเหล่านี้ทำให้ผู้ตอบรู้สึกไม่ปลอดภัยหากตนตอบว่าเสพ ดังนั้นโอกาสที่เรจะได้ข้อมูลที่แท้จริงเป็นสิ่งที่ทำได้ยาก วอร์เนอร์จึงคิดวิธีการหรือเครื่องมือโดยในคำถามปกปิดหนึ่ง ๆ แบ่งออกเป็นสองลักษณะคือลักษณะเชิงบวก (Q_A) และลักษณะเชิงนิเสธของคำถามแรก (Q_B) ดังนี้

Q_A : “ท่านเป็นสมาชิกในกลุ่ม A ใช่หรือไม่”

Q_B : “ท่านไม่ได้เป็นสมาชิกในกลุ่ม A ใช่หรือไม่”

เช่น Q_A : “ท่านเคยเสพยาบ้า ใช่หรือไม่”

Q_B : “ท่านไม่เคยเสพยาบ้าหรือไม่”

ก่อนที่จะตอบคำถามให้ผู้ตอบสุ่มหยิบคำถามแบบใดหนึ่งในเครื่องมือเพียงหนึ่งคำถาม เช่น ให้ผู้ถูกสัมภาษณ์ เลือกคำถาม จากไฟ 1 ชุด ซึ่งประกอบด้วย ไฟสีแดง และไฟสีดำ ในอัตราส่วนที่ผู้สำรวจได้กำหนดไว้ล่วงหน้าแล้ว ในการตอบคำถามผู้สัมภาษณ์จะให้ผู้ถูกสัมภาษณ์สับไฟจนเข้ากันดีเพื่อให้ไฟแต่ละใบมีโอกาสถูกสุ่มด้วยความน่าจะเป็นเท่า ๆ กันแล้วจึงมา 1 ใบจากชุดนั้นโดยไม่ให้ผู้อื่นทราบว่าได้ไฟสีอะไร ถ้าผู้สัมภาษณ์ดึงได้ไฟสีแดงจะต้องตอบคำถามข้อแรกซึ่งเกี่ยวกับคำถามที่ปกปิดที่ต้องการศึกษา เช่น “ท่านเคยเสพยาบ้า ใช่หรือไม่” แต่ถ้าผู้ถูกสัมภาษณ์ดึงได้ไฟสีดำจะต้องตอบคำถามข้อสองซึ่งเป็นนิเสธของคำถามแรก เช่น “ท่านไม่เคยเสพยาบ้า ใช่หรือไม่” เมื่อผู้ถูกสัมภาษณ์ทุกคนเข้าใจวิธีการ และตอบคำถามตามความจริงแล้วประมาณค่าสัดส่วนของประชากรที่มีลักษณะปกปิด (π) ด้วยวิธีภาวะน่าจะเป็นสูงสุด และเขาได้ตั้งชื่อเทคนิคว่า เทคนิคการตอบสนองการสุ่ม ในการประมาณสัดส่วนที่แท้จริงได้จำแนกออกเป็น 2 กรณีดังนี้

1) กรณีที่ผู้ถูกสัมภาษณ์ทุกคนตอบตามความจริง (Completely truthful)

กำหนดให้	π	แทนสัดส่วนที่แท้จริงของประชากรที่มีลักษณะปกปิด
	π_w	แทนตัวประมาณสัดส่วนที่แท้จริงของประชากรที่มีลักษณะปกปิดโดยวิธี วอร์เนอร์กรณีผู้ถูกสัมภาษณ์ทุกคนตอบตามความจริง
	p	แทนความน่าจะเป็นที่จะสุ่มได้คำถามที่ปกปิดเชิงบวก
X	$= \begin{cases} 1 & \text{ถ้าผู้ถูกสัมภาษณ์ ตอบว่า "ใช่"} \\ 0 & \text{ถ้าผู้ถูกสัมภาษณ์ ตอบว่า "ไม่ใช่"} \end{cases}$	

ดังนั้น $P(X=1) = \pi p + (1-\pi)(1-p)$

$P(X=0) = 1 - \{\pi p + (1-\pi)(1-p)\} = (1-\pi)p + \pi(1-p)$

ให้	n_1	แทนจำนวนผู้ถูกสัมภาษณ์ที่ตอบว่า "ใช่" จากจำนวนผู้ถูกสัมภาษณ์ n คน
	$n - n_1$	แทนจำนวนผู้ถูกสัมภาษณ์ที่ตอบว่า "ไม่ใช่"

จะได้ ฟังก์ชันภาวะน่าจะเป็น

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

ด้วยวิธีการหาตัวประมาณแบบภาวะน่าจะเป็นสูงสุด หาตัวประมาณได้ดังนี้

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

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

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

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

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

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

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

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

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

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

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

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

นั่นคือ ค่าประมาณภาวะน่าจะเป็นสูงสุด ของ π คือ

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

ซึ่งเป็นตัวประมาณที่ไม่เอนเอียงดังแสดงต่อไปนี้

$$\begin{aligned}
 E(\hat{\pi}_w) &= \frac{p-1}{2p-1} + \frac{E(n_1)}{(2p-1)n} \\
 &= \frac{1}{2p-1} \left[p-1 + (1/n) \sum_{i=1}^n EX_i \right] \\
 &= \frac{1}{2p-1} \left[p-1 + (1/n) \sum_{i=1}^n EX_i \right] \\
 &= \frac{1}{2p-1} [p-1 + \pi p + (1-\pi)(1-p)] \\
 &= \frac{1}{2p-1} [p-1 + \pi p + 1 - p - \pi + \pi p] \\
 &= \pi
 \end{aligned}$$

และมีความแปรปรวนของ $\hat{\pi}_w$ ดังนี้

$$\begin{aligned}
 \text{Var}(\hat{\pi}_w) &= \text{Var} \left[\frac{p-1}{2p-1} + \frac{n_1}{(2p-1)n} \right] \\
 &= \text{Var} \left[\frac{n_1}{(2p-1)n} \right] \\
 &= \frac{\text{Var}(n_1)}{(2p-1)^2 n^2}
 \end{aligned}$$

$$\begin{aligned}
&= \frac{[\pi p + (1-\pi)(1-p)] [(1-\pi)p + \pi(1-p)]}{(2p-1)^2 n} \\
&= \frac{1/4 + (2p^2 - 2p + 1/2)(-2\pi^2 + 2\pi - 1/2)}{(2p-1)^2 n} \\
&= \frac{1}{n} \left[\frac{1}{16(p-1/2)^2} - (\pi - 1/2)^2 \right]
\end{aligned}$$

$$\text{หรือ } \text{Var}(\hat{\pi}_w) = \frac{1}{4n} - \frac{(\pi - 1/2)^2}{n} + \frac{1}{16(p-1/2)^2 n} - \frac{1}{4n}; \quad p \neq \frac{1}{2}$$

หรือสามารถเขียน ความแปรปรวนของ ในรูป

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

จากสมการข้างต้น ความแปรปรวนของ $\hat{\pi}_w$ สามารถเขียนในรูปของผลรวมระหว่าง ความแปรปรวนเนื่องจากการสุ่มตัวอย่าง (Variance due to sampling) และความแปรปรวนเนื่องจากเครื่องสุ่มเลือก (Variance due to randomizing device)

2) กรณีที่บางส่วนของผู้ถูกสัมภาษณ์ที่มีลักษณะปกปิด ตอบไม่ตรงตามความจริง

(Less than completely truthful)

ให้ π'_w แทน ตัวประมาณสัดส่วนที่แท้จริงของประชากรที่มีลักษณะปกปิดโดยวิธีวอร์เนอร์
กรณีที่บางส่วนของผู้ถูกสัมภาษณ์ที่มีลักษณะปกปิด ตอบไม่ตรงตามความจริง

T_w แทน ความน่าจะเป็นที่ผู้ถูกสัมภาษณ์ที่มีลักษณะปกปิดตอบคำถามปกปิดตามความจริง โดยที่ $0 < T_w \leq 1$

ดังนั้น

$$P(X=1) = \pi p T_w + \pi(1-p)(1-T_w) + (1-\pi)(1-p)$$

$$P(X=0) = 1 - \{ \pi p T_w + \pi(1-p)(1-T_w) + (1-\pi)(1-p) \}$$

ให้ n_2 แทน จำนวนผู้ถูกสัมภาษณ์ที่ตอบว่า “ใช่” จากจำนวนผู้ถูกสัมภาษณ์ทั้งหมด n คน

ค่าประมาณของ π หาได้ดังนี้

ฟังก์ชันภาษาน่าจะเป็น

$$L(\pi) = [\pi p T_w + \pi(1-p)(1-T_w) + (1-\pi)(1-p)]^{n_2} [1 - \{\pi p T_w + \pi(1-p)(1-T_w) + (1-\pi)(1-p)\}]^{n-n_2}$$

ด้วยวิธีการหาตัวประมาณแบบภาษาน่าจะเป็นสูงสุด หาตัวประมาณได้

$$\begin{aligned} \log L(\pi) &= n_2 \log [\pi p T_w + \pi(1-p)(1-T_w) + (1-\pi)(1-p)] + \\ &\quad (n - n_2) \log [1 - \{\pi p T_w + \pi(1-p)(1-T_w) + (1-\pi)(1-p)\}] \\ \frac{d}{d\pi} \log L(\pi) &= \frac{n_2 [p T_w + (1-p)(1-T_w) - (1-p)]}{\pi p T_w + \pi(1-p)(1-T_w) + (1-\pi)(1-p)} + \frac{(n - n_2) [-p T_w - (1-p)(1-T_w) + (1-p)]}{1 - \{\pi p T_w + \pi(1-p)(1-T_w) + (1-\pi)(1-p)\}} \\ &= \frac{n_2 (2p T_w - T_w)}{\pi p T_w + \pi(1-p)(1-T_w) + (1-\pi)(1-p)} - \frac{(n - n_2) (2p T_w - T_w)}{[1 - \{\pi p T_w + \pi(1-p)(1-T_w) + (1-\pi)(1-p)\}]} \\ &\stackrel{0}{=} \frac{n_2 (2p T_w - T_w)}{\pi p T_w + \pi(1-p)(1-T_w) + (1-\pi)(1-p)} - \frac{(n - n_2) (2p T_w - T_w)}{[1 - \{\pi p T_w + \pi(1-p)(1-T_w) + (1-\pi)(1-p)\}]} \\ &= \frac{n_2 (2p T_w - T_w)}{\pi p T_w + \pi(1-p)(1-T_w) + (1-\pi)(1-p)} = \frac{(n - n_2) (2p T_w - T_w)}{[1 - \{\pi p T_w + \pi(1-p)(1-T_w) + (1-\pi)(1-p)\}]} \\ &\quad \frac{[1 - \{\pi p T_w + \pi(1-p)(1-T_w) + (1-\pi)(1-p)\}]}{\pi p T_w + \pi(1-p)(1-T_w) + (1-\pi)(1-p)} = \frac{n - n_2}{n_2} \end{aligned}$$

$$\frac{\left[1 - \left\{ \pi_p T_w + \pi(1-p)(1-T_w) + (1-\pi)(1-p) \right\}\right]}{\pi_p T_w + \pi(1-p)(1-T_w) + (1-\pi)(1-p)} = \frac{n}{n_2} - 1$$

$$\frac{\left[1 - \left\{ \pi_p T_w + \pi(1-p)(1-T_w) + (1-\pi)(1-p) \right\}\right]}{\pi_p T_w + \pi(1-p)(1-T_w) + (1-\pi)(1-p)} + 1 = \frac{n}{n_2}$$

$$\frac{1}{\pi_p T_w + \pi(1-p)(1-T_w) + (1-\pi)(1-p)} = \frac{n}{n_2}$$

$$\pi_p T_w + \pi(1-p)(1-T_w) + (1-\pi)(1-p) = \frac{n_2}{n}$$

$$\pi_p T_w + (1-p) \left[\pi(1-T_w) + 1 - \pi \right] = \frac{n_2}{n}$$

$$\pi T_w (2p-1) - (p-1) = \frac{n_2}{n}$$

$$\pi T_w = \frac{(p-1)}{(2p-1)} + \frac{n_2}{(2p-1)n}$$

นั่นคือ ค่าประมาณภาวะน่าจะเป็นสูงสุด ของ π กรณีบางส่วนของผู้ถูกสัมภาษณ์ที่มีลักษณะปกปิด
ตอบไม่ตรงตามความจริง เท่ากับ

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

เมื่อ

$$\hat{\pi}'_w = \hat{\pi}_w T_w$$

ค่าคาดหวัง (Expected value) ของ ค่าประมาณ คือ

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

นั่นคือ $\hat{\pi}'_w$ เป็นตัวประมาณที่เอนเอียงของ π

ดังนั้นค่าความเอนเอียงของ $\hat{\pi}'_w$ คือ

$$\begin{aligned} \text{Bias}(\hat{\pi}'_w) &= E(\hat{\pi}'_w) - \pi \\ &= \frac{p-1}{2p-1} + \frac{\pi p T_w + \pi(1-p)(1-T_w) + (1-\pi)(1-p)}{(2p-1)} - \pi \\ &= \frac{p-1 + \pi p T_w + \pi(1-p)(1-T_w) + (1-p)(1-\pi) - \pi(2p-1)}{(2p-1)} \\ &= \frac{\pi[p T_w + (1-p)(1-T_w) - 2p + 1] + (1-\pi)(1-p) - (1-p)}{(2p-1)} \\ &= \frac{\pi[2p T_w - T_w + 2 - 3p] + \pi(p-1)}{(2p-1)} \\ &= \frac{\pi[(2p-1)T_w - (2p+1)]}{(2p-1)} \\ &= \frac{\pi(2p-1)(T_w - 1)}{(2p-1)} \\ &= \pi(T_w - 1) \end{aligned}$$

ความแปรปรวนของ $\hat{\pi}'_w$ คือ

$$\begin{aligned}
 \text{Var}(\hat{\pi}'_w) &= \text{Var}\left[\frac{p-1}{2p-1} + \frac{n_2}{(2p-1)n}\right] \\
 &= \frac{\text{Var}(n_2)}{(2p-1)^2 n^2} \\
 &= \frac{n}{(2p-1)^2 n^2} [\pi p T_w + \pi(1-p)(1-T_w) + (1-\pi)(1-p)] \cdot \\
 &\quad [1 - \{\pi p T_w + \pi(1-p)(1-T_w) + (1-\pi)(1-p)\}] \\
 &= \frac{[\pi T_w(2p-1) + (1-p)] \cdot [p - \pi T_w(2p-1)]}{(2p-1)^2 n} \\
 &= \frac{\pi T_w(2p-1)p - \pi^2 T_w^2(2p-1)^2 + (1-p)p - (1-p)(2p-1)\pi T_w}{(2p-1)^2 n} \\
 &= \frac{\pi T_w(2p-1)^2 - \pi^2 T_w^2(2p-1)^2 + (1-p)p}{(2p-1)^2 n} \\
 &= \frac{\pi T_w(2p-1)^2 [1 - \pi T_w]}{(2p-1)^2 n} + \frac{(1-p)p}{(2p-1)^2 n}
 \end{aligned}$$

นั่นคือ $\text{Var}(\hat{\pi}'_w) = \frac{\pi T_w(1 - \pi T_w)}{n} + \frac{p(1-p)}{(2p-1)^2 n}; p \neq \frac{1}{2}$

2.4.2 กรีนเบิร์กและคณะ [23]

หลังจากวอร์เนอร์ได้คิดเทคนิคการตอบสนองการสุ่มแล้วซิมมอนส์ (Simmons) [อ้างใน [23]] ได้ดัดแปลงวิธีวอร์เนอร์โดย เขามีแนวคิดว่าในเครื่องมือ มีคำถามปกปิด (Q_A) แล้วอีกคำถามหนึ่งเป็นคำถามที่ไม่เกี่ยวข้องกับเรื่องที่มีลักษณะปกปิดที่ต้องการศึกษาแต่เป็นคำถามทั่ว ๆ ไปหรือคำถามที่ไม่ปกปิด (Q_Y) เป็นอีกตัวเลือกจะทำให้ผู้ตอบกล้าที่จะตอบความจริงมากกว่าที่จะใช้คำถามเชิงนิเสธที่มีลักษณะปกปิดที่เกี่ยวข้องกัน ซึ่งจะทำให้ลดความแปรปรวนของตัวประมาณด้วย ลักษณะคำถามเป็นดังนี้

Q_A : “ท่านเป็นสมาชิกในกลุ่ม A ใช่หรือไม่”

Q_Y : “ท่านเป็นสมาชิกในกลุ่ม Y ใช่หรือไม่”

เช่น Q_A : “ท่านเคยเสพยาบ้าหรือไม่”

Q_Y : “ท่านเกิดเดือนเมษายนใช่หรือไม่”

ต่อมากรีนเบิร์กและคณะได้คิดค้นวิธีการโดยดัดแปลงจากแนวความคิดของซิมมอนส์ เรียกเทคนิคนี้ว่า Alternate Question Randomized Response Techniques หรือ Unrelated Question Randomized Techniques โดยเครื่องมือในการเก็บรวบรวมข้อมูลเป็นไปในลักษณะที่คล้ายกับวอร์เนอร์จะต่างกันที่ คำถามในเครื่องมือ ดังที่กล่าวไว้แล้วข้างต้น แต่ในการประมาณค่าสัดส่วนในครั้งนี้ เราไม่เพียงประมาณสัดส่วนของประชากรที่มีลักษณะปกปิด (π) เท่านั้นเราจำเป็นต้องหาตัวประมาณสัดส่วนของประชากรที่มีลักษณะไม่ปกปิด (π_Y) ด้วย จึงจะทำการแก้สมการเพื่อหา π ได้จึงจำเป็นต้องกำหนด ตัวอย่าง 2 กลุ่มที่เป็นอิสระต่อกัน แต่ถ้าทราบสัดส่วนของประชากรที่มีลักษณะโดยทั่วๆ ไปซึ่งไม่จำเป็นต้องปกปิด (π_Y) ล่วงหน้า จะสามารถลดตัวอย่างโดยใช้ตัวอย่างเพียงกลุ่มเดียวได้ ดังนั้นในการศึกษาวิธีการจึงแบ่งเป็น 2 กรณีดังนี้

1) กรณีที่ไม่ทราบสัดส่วนของประชากรที่มีลักษณะไม่ปกปิด

สุ่มตัวอย่างมา 2 กลุ่ม โดยสมาชิกแต่ละกลุ่มไม่จำเป็นต้องเท่ากันและมีเครื่องมือการตอบสนองการสุ่ม 2 ชุด ในเครื่องมือชุดที่ 1 และ 2 ประกอบด้วยคำถาม 2 คำถาม คือ Q_A และ Q_Y โอกาสที่หน่วยตัวอย่างที่ 1 จะสุ่มหยิบได้คำถาม Q_A จากเครื่องมือชุดที่ 1 ด้วยความน่าจะเป็น p_1 และโอกาสที่หน่วยตัวอย่างที่ 2 จะสุ่มหยิบได้คำถาม Q_A จากเครื่องมือชุดที่ 2 ด้วยความน่าจะเป็น p_2 และทำนองกลับกันโอกาสที่ตัวอย่างที่ 1 จะสุ่มหยิบได้คำถาม Q_Y จากเครื่องมือชุดที่ 1 ด้วยความน่าจะเป็น $1-p_1$ และ โอกาสที่ตัวอย่างที่ 2 จะสุ่มหยิบได้คำถาม Q_Y จากเครื่องมือชุดที่ 2 ด้วยความน่าจะเป็น $1-p_2$ ด้วยวิธีการเก็บรวบรวมข้อมูล ก่อนที่จะตอบคำถามให้ผู้ตอบสุ่มหยิบคำถามแบบใส่คืน เพียงหนึ่งคำถาม เช่นให้ผู้ถูกสัมภาษณ์เลือกคำถามจากไฟที่ประกอบด้วย ไฟสีแดงและสีดำในอัตราส่วนที่ผู้สำรวจกำหนดไว้ล่วงหน้าแล้ว ในการตอบคำถามผู้สัมภาษณ์จะให้

ผู้ถูกสัมภาษณ์สลับไปจนเข้ากันดี เพื่อให้ไฟแต่ละใบมีโอกาสถูกสุ่มด้วยความน่าจะเป็นเท่า ๆ กัน แล้วดึงมา 1 ใบจากชุดนั้นโดยไม่ให้ผู้อื่นทราบว่าได้ไฟสีอะไร ถ้าผู้สัมภาษณ์ดึงได้ไฟสีแดงจะต้องตอบคำถามข้อแรกซึ่งเกี่ยวกับคำถามที่ปกปิดที่ต้องการศึกษา (Q_A) แต่ถ้าผู้ถูกสัมภาษณ์ดึงได้ไฟสีดำ จะต้องตอบคำถามที่ไม่เกี่ยวข้องกับคำถามที่ปกปิดที่ต้องการศึกษา (Q_Y) ว่า “ใช่” หรือ “ไม่ใช่” โดยตัวอย่างที่ 1 สุ่มหยิบไฟจากชุดที่ 1 ส่วนตัวอย่างที่ 2 สุ่มหยิบไฟจากชุดที่ 2 เมื่อผู้ถูกสัมภาษณ์ทุกคนเข้าใจวิธีการ และตอบคำถามตามความจริงแล้ว สามารถประมาณค่าสัดส่วนของประชากรที่มีลักษณะปกปิด (π) โดยใช้สถิติและความน่าจะเป็นดังต่อไปนี้

1.1) กรณีที่ผู้ถูกสัมภาษณ์ทุกคนตอบความจริง

กำหนดให้ $(\hat{\pi}_1)_U$ แทน	ตัวประมาณสัดส่วนที่แท้จริงของประชากรที่มีลักษณะปกปิดโดยวิธีกรีนเบิร์กและคณะ กรณีที่ไม่ทราบสัดส่วนของประชากรที่มีลักษณะไม่ปกปิดและผู้ถูกสัมภาษณ์ทุกคนตอบตามความจริง
$(\hat{\pi})_Y$ แทน	ตัวประมาณสัดส่วนที่แท้จริงของประชากรที่มีลักษณะไม่ปกปิด
p_i แทน	ความน่าจะเป็นที่ตัวอย่าง i จะสุ่มได้คำถามที่ปกปิด เมื่อ $i = 1, 2$
n แทน	จำนวนผู้ถูกสัมภาษณ์ทั้งหมด ซึ่ง $n = n_1 + n_2$ เมื่อ n_1, n_2 แทนขนาดตัวอย่าง 1 และ 2 ตามลำดับ
X_i แทน	จำนวนผู้ถูกสัมภาษณ์ที่ตอบว่าใช่จากตัวอย่างที่ i เมื่อ $i = 1, 2$
λ_i แทน	ความน่าจะเป็นที่ผู้ตอบจะตอบว่าใช่จากตัวอย่าง i เมื่อ $i = 1, 2$

เนื่องจาก

$$\lambda_1 = p_1 \pi + (1 - p_1) \pi_Y \quad (2.1)$$

$$\lambda_2 = p_2 \pi + (1 - p_2) \pi_Y \quad (2.2)$$

ต้องการหา π, π_Y โดยการแก้สมการ (2.1), (2.2)

$$\text{จาก(2.1); } \lambda_1(1 - p_2) = p_1 \pi(1 - p_2) + (1 - p_1)(1 - p_2) \pi_Y \quad (2.3)$$

$$\text{จาก(2.2); } \lambda_2(1 - p_1) = p_2 \pi(1 - p_1) + (1 - p_2)(1 - p_1) \pi_Y \quad (2.4)$$

$$(2.3)-(2.4); \lambda_1(1 - p_2) - \lambda_2(1 - p_1) = \pi[p_1(1 - p_2) - p_2(1 - p_1)]$$

$$\lambda_1(1-p_2) - \lambda_2(1-p_1) = \pi(p_1 - p_2)$$

$$\pi = \frac{\lambda_1(1-p_2) - \lambda_2(1-p_1)}{p_1 - p_2} \quad \text{เมื่อ } p_1 \neq p_2$$

$$\text{จาก(2.1); } \lambda_{1p_2} = p_1 p_2 \pi + (1-p_1)p_2 \pi_Y \quad (2.5)$$

$$\text{จาก(2.2); } \lambda_{2p_1} = p_1 p_2 \pi + (1-p_2)p_1 \pi_Y \quad (2.6)$$

$$(2.5)-(2.6); \lambda_{1p_2} - \lambda_{2p_1} = \pi_Y [(1-p_1)p_2 - (1-p_2)p_1]$$

$$\lambda_{1p_2} - \lambda_{2p_1} = \pi_Y (p_2 - p_1)$$

$$\pi_Y = \frac{\lambda_{1p_2} - \lambda_{2p_1}}{p_2 - p_1} \quad \text{เมื่อ } p_2 \neq p_1$$

ให้ $\hat{\lambda}_1, \hat{\lambda}_2$ แทน สัดส่วนของผู้ที่ตอบว่าใช่จากตัวอย่าง 1 และ 2 ตามลำดับและ X_1, X_2 มีการแจกแจงแบบทวินามด้วยพารามิเตอร์ $(n_1, \lambda_1), (n_2, \lambda_2)$ ตามลำดับซึ่ง

$$\hat{\lambda}_1 = \frac{X_1}{n_1} \quad \text{และ} \quad \hat{\lambda}_2 = \frac{X_2}{n_2}$$

ดังนั้นตัวประมาณของ π, π_Y คือ

$$(\hat{\pi})_u = \frac{\hat{\lambda}_1(1-p_2) - \hat{\lambda}_2(1-p_1)}{p_1 - p_2} \quad \text{เมื่อ } p_2 \neq p_1$$

$$(\hat{\pi})_Y = \frac{\hat{\lambda}_{1p_2} - \hat{\lambda}_{2p_1}}{p_2 - p_1} \quad \text{เมื่อ } p_2 \neq p_1$$

ซึ่ง $(\hat{\pi}_1)_u$ และ $(\hat{\pi})_y$ ต่างก็เป็นตัวประมาณที่ไม่เอนเอียงของ π, π_y ตามลำดับดังแสดงต่อไปนี้

$$\begin{aligned} E(\hat{\pi}_1)_u &= \frac{E(\hat{\lambda}_1)(1-p_2) - E(\hat{\lambda}_2)(1-p_1)}{p_1 - p_2} \\ &= \frac{\lambda_1(1-p_2) - \lambda_2(1-p_1)}{p_1 - p_2} \\ &= \frac{[p_1\pi + (1-p_1)\pi_y](1-p_2) - [p_2\pi + (1-p_2)\pi_y](1-p_1)}{p_1 - p_2} \\ &= \frac{\pi[p_1(1-p_2) - p_2(1-p_1)]}{p_1 - p_2} \end{aligned}$$

และ

$$\begin{aligned} E(\hat{\pi})_y &= \frac{E(\hat{\lambda}_1)p_2 - E(\hat{\lambda}_2)p_1}{p_2 - p_1} \\ &= \frac{\lambda_1 p_2 - \lambda_2 p_1}{p_2 - p_1} \\ &= \frac{[p_1\pi + (1-p_1)\pi_y](p_2) - [p_2\pi + (1-p_2)\pi_y](p_1)}{p_2 - p_1} \\ &= \pi_y \end{aligned}$$

และสามารถหาความแปรปรวนของตัวประมาณได้ดังนี้

$$\begin{aligned}\text{Var}(\hat{\pi}_1)_u &= \text{Var}\left(\frac{\hat{\lambda}_1(1-p_2) - \hat{\lambda}_2(1-p_1)}{p_1 - p_2}\right) \\ &= \frac{(1-p_2)^2 \text{Var}(\hat{\lambda}_1) + (1-p_1)^2 \text{Var}(\hat{\lambda}_2)}{(p_1 - p_2)^2} \\ &= \frac{1}{(p_1 - p_2)^2} \left\{ \frac{\lambda_1(1-\lambda_1)(1-p_2)^2}{n_1} + \frac{\lambda_2(1-\lambda_2)(1-p_1)^2}{n_2} \right\}\end{aligned}$$

นั่นคือความแปรปรวนของตัวประมาณสัดส่วนด้วยวิธีการของกรีนเบิร์กและคณะ จะมีค่าน้อยเมื่อ $p_1 - p_2$ มีค่ามาก และในขณะเดียวกันอยู่กับ ค่า n_1, n_2 ด้วยซึ่งถ้า

$$\frac{\lambda_1(1-\lambda_1)(1-p_2)^2}{n_1} \text{ และ } \frac{\lambda_2(1-\lambda_2)(1-p_1)^2}{n_2} \text{ ทำให้ได้ สัดส่วนของขนาดตัวอย่างที่เหมาะสมคือ}$$

$$\frac{n_1}{n_2} = \left(\frac{\lambda_1(1-\lambda_1)(1-p_2)^2}{\lambda_2(1-\lambda_2)(1-p_1)^2} \right)^{1/2}$$

ซึ่งจะเป็นสัดส่วนที่ทำให้แต่ละตัวอย่างมีโอกาสที่จะได้ตอบคำถาม Q_A มากกว่า Q_Y นั้นเอง

1.2) กรณีที่บางส่วนของผู้ถูกสัมภาษณ์ที่มีลักษณะปกปิด ตอบไม่ตรงตามความจริง

กำหนดให้

$(\hat{\pi}_1')_u$ แทน ตัวประมาณสัดส่วนที่แท้จริงของประชากรที่มีลักษณะปกปิดโดยวิธีการของกรีนเบิร์กและคณะ กรณีที่ไม่ทราบสัดส่วนของประชากรที่มีลักษณะไม่ปกปิดเมื่อบางส่วนของผู้ถูกสัมภาษณ์ที่มีลักษณะปกปิดตอบคำถามไม่ตรงตามความจริง

T_U แทน ความน่าจะเป็นที่ผู้ถูกสัมภาษณ์ที่มีลักษณะปกปิดตอบคำถามไม่ตรงตามความจริงและผู้ที่มีลักษณะไม่ปกปิดตอบคำถามปกปิดตามความจริงโดยที่ $0 < T_U \leq 1$

- $(\hat{\pi}'_i)_Y$ แทน ตัวประมาณสัดส่วนที่แท้จริงของประชากรที่มีลักษณะไม่ปกปิดกรณีที่ไม่ทราบสัดส่วนของประชากรที่มีลักษณะไม่ปกปิดเมื่อบางส่วนของผู้ถูกสัมภาษณ์ที่มีลักษณะปกปิดตอบคำถามไม่ตรงตามความจริง
- Y_i แทน จำนวนผู้ถูกสัมภาษณ์ที่ตอบว่าใช่จากตัวอย่างที่ i เมื่อ $i = 1, 2$ จากผู้ถูกสัมภาษณ์กลุ่ม 1 และ 2 เป็น n_1, n_2 คน ตามลำดับ
- λ'_i แทน ความน่าจะเป็นที่ผู้ตอบจะตอบว่าใช่จากตัวอย่าง i เมื่อ $i = 1, 2$ กรณีที่บางส่วนของผู้ถูกสัมภาษณ์ที่มีลักษณะปกปิด ตอบไม่ตรงตามความจริง

ดังนั้น

$$\lambda'_1 = p_1 \pi_{T_U} (1 - p_1) + \pi_Y \quad (2.7)$$

$$\lambda'_2 = p_2 \pi_{T_U} (1 - p_2) + \pi_Y \quad (2.8)$$

ต้องการหา π, π_Y โดยการแก้สมการ (2.7), (2.8)

จาก(2.7); $\lambda'_1 (1 - p_2) = p_1 \pi_{T_U} (1 - p_2) (1 - p_1) + (1 - p_2) \pi_Y \quad (2.9)$

จาก(2.8); $\lambda'_2 (1 - p_1) = p_2 \pi_{T_U} (1 - p_1) (1 - p_2) + (1 - p_1) \pi_Y \quad (2.10)$

(2.9)-(2.10); $\lambda'_1 (1 - p_2) - \lambda'_2 (1 - p_1) = \pi_Y [(1 - p_2) - (1 - p_1)]$

$$\lambda'_1 (1 - p_2) - \lambda'_2 (1 - p_1) = \pi_Y (p_1 - p_2)$$

$$\pi_Y = \frac{\lambda'_1 (1 - p_2) - \lambda'_2 (1 - p_1)}{p_1 - p_2} \quad \text{เมื่อ } p_1 \neq p_2$$

จาก(2.7); $\lambda'_{1p_2} = p_1 p_2 (1 - p_1) \pi_{T_U} + p_2 \pi_Y \quad (2.11)$

จาก(2.8); $\lambda'_{2p_1} = p_1 p_2 (1 - p_2) \pi_{T_U} + p_1 \pi_Y \quad (2.12)$

(2.11)-(2.12); $\lambda'_{1p_2} - \lambda'_{2p_1} = \pi_Y [(1 - p_1) p_2 - (1 - p_2) p_1]$

$$\lambda'_1 p_2 - \lambda'_2 p_1 = \pi_Y (p_2 - p_1)$$

$$\pi_Y = \frac{\lambda'_1 p_2 - \lambda'_2 p_1}{p_2 - p_1} \quad \text{เมื่อ } p_2 \neq p_1$$

ให้ $\hat{\lambda}'_1, \hat{\lambda}'_2$ แทน สัดส่วนของผู้ที่ตอบว่าใช่จากตัวอย่าง 1 และ 2 ตามลำดับและ Y_1, Y_2 มีการแจกแจงแบบทวินามด้วยพารามิเตอร์ $(n_1, \lambda'_1), (n_2, \lambda'_2)$ ตามลำดับโดยที่

$$\hat{\lambda}_1 = \frac{Y_1}{n_1} \quad \text{และ} \quad \hat{\lambda}_2 = \frac{Y_2}{n_2}$$

ดังนั้นตัวประมาณของ π, π_Y คือ

$$\begin{aligned} (\hat{\pi}'_1)_u &= \frac{\hat{\lambda}'_1(1-p_2) - \hat{\lambda}'_2(1-p_1)}{p_1 - p_2} & \text{เมื่อ } p_2 \neq p_1 \\ (\hat{\pi}'_1)_Y &= \frac{\hat{\lambda}'_1(1-p_2) - \hat{\lambda}'_2(1-p_1)}{p_1 - p_2} & \text{เมื่อ } p_2 \neq p_1 \end{aligned}$$

ค่าคาดหวังของ $(\hat{\pi}'_1)_u$ คือ

$$E(\hat{\pi}'_1)_u = \frac{\lambda'_1(1-p_2) - \lambda'_2(1-p_1)}{p_1 - p_2}$$

นั่นคือ $(\hat{\pi}'_1)_u$ เป็นตัวประมาณที่เอนเอียงของ π

ดังนั้น ค่าความเอนเอียง และความแปรปรวนของ $(\hat{\pi}'_1)_u$ พิจารณาได้ดังนี้

$$\text{Bias}(\hat{\pi}'_1)_u = E(\hat{\pi}'_1 - \pi)_u$$

$$\begin{aligned} &= \frac{(\lambda'_1 - \lambda_1)(1-p_2) - (\lambda'_2 - \lambda_2)(1-p_1)}{p_1 - p_2} \end{aligned}$$

$$\begin{aligned}
&= \frac{\left\{ \left[p_1 \pi_{T_U} + (1-p_1) \pi_Y \right] - \left[p_1 \pi + (1-p_1) \pi_Y \right] \right\} (1-p_2)}{p_1 - p_2} \\
&\quad - \frac{\left\{ \left[p_2 \pi_{T_U} + (1-p_2) \pi_Y \right] - \left[p_2 \pi + (1-p_2) \pi_Y \right] \right\} (1-p_2)}{p_1 - p_2} \\
&= \frac{\pi_{T_U} (p_1 (1-p_2) - p_2 (1-p_1))}{p_1 - p_2} \\
&= \pi(T_U) \\
\text{Var}(\hat{\pi}'_1)_u &= \frac{(1-p_2)^2 \text{Var}(\hat{\lambda}'_1) + (1-p_1)^2 \text{Var}(\hat{\lambda}'_2)}{(p_1 - p_2)^2} \\
&= \frac{1}{(p_1 - p_2)^2} \left\{ \frac{\lambda'_1 (1-\lambda'_1) (1-p_2)^2}{n_1} + \frac{\lambda'_2 (1-\lambda'_2) (1-p_1)^2}{n_2} \right\}
\end{aligned}$$

2) กรณีที่ทราบสัดส่วนของประชากรที่มีลักษณะไม่ปกปิด

โดยบางกรณีเราอาจทราบถึงลักษณะของประชากรโดยทั่วไป เช่น เมื่อ π_Y คือ สัดส่วนของคนที่เกิดในเดือนเมษายน เราสามารถทราบข้อมูลเบื้องต้นนี้จากสำมะโนประชากร หรือทะเบียนราษฎรได้ นั่นคือ ถ้าเราทราบ π_Y ตัวประมาณที่ต้องการหาจะมีเพียงค่าเดียวดังนั้นใช้เพียงตัวอย่างเดียวก็เพียงพอแล้วสามารถหาตัวประมาณได้ 2 กรณีดังนี้

2.1) กรณีที่ผู้ถูกสัมภาษณ์ทุกคนตอบความจริง

กำหนดให้ p แทน ความน่าจะเป็นที่ผู้ตอบจะสุ่มได้คำถามที่มีลักษณะปกปิด

$(\pi_2 | \pi_Y)_u$ แทน สัดส่วนที่แท้จริงของประชากรที่มีลักษณะปกปิดโดยที่ทราบสัดส่วนของประชากรที่มีลักษณะไม่ปกปิดและทุกคนตอบตามความจริง

$(\hat{\pi}_2 | \pi_Y)_u$ แทน สัดส่วนของตัวอย่างที่มีลักษณะปกปิดโดยที่ทราบสัดส่วนของประชากรที่มีลักษณะไม่ปกปิดและทุกคนตอบตามความจริง

λ แทน ความน่าจะเป็นที่ผู้ตอบจะตอบว่า “ใช่”

ดังนั้น $\lambda = p(\pi - \pi_Y) + \pi_Y$

หรือ $\lambda = p\pi + (1-p)\pi_Y$

และจะได้
$$\pi = \frac{\lambda - \pi_Y(1-p)}{p}$$

ให้ $\hat{\lambda}$ แทนสัดส่วนของผู้ที่ตอบว่าใช่จากตัวอย่างและ n_1 มีการแจกแจงแบบทวินามด้วยพารามิเตอร์ (n, λ) เมื่อ n แทนขนาดตัวอย่าง และ n_1 แทนจำนวนผู้ที่ตอบว่า “ใช่”, $\hat{\lambda} = \frac{n_1}{n}$

ดังนั้นตัวประมาณคือ

$$(\hat{\pi}_2 | \pi_Y)_U = \frac{\hat{\lambda} - \pi_Y(1-p)}{p}$$

ซึ่งเป็นตัวประมาณที่ไม่เอนเอียง

สามารถหาความแปรปรวนของตัวประมาณได้คือ

$$\text{Var}(\hat{\pi}_2 | \pi_Y)_U = \frac{1}{np^2} \{p\pi + (1-p)\pi_Y\} \{1 - p\pi - (1-p)\pi_Y\}$$

2.2) กรณีที่บางส่วนของผู้ถูกสัมภาษณ์ที่มีลักษณะปกปิด ตอบไม่ตรงตามความจริง

ให้ π_Y^* แทนค่าประมาณของ π_Y ที่ใช้ในวิธีของกรีนเบิร์กและคณะ ซึ่ง π_Y^* จะมากกว่าค่าที่ได้จากการสำรวจ นั่นคือ

$$\pi_Y = \pi_Y^* - C \quad \text{เมื่อ } C \text{ เป็นค่าคงตัว}$$

ในกรณีนี้สามารถประมาณ $(\pi | \pi_Y)_U$ ได้คือ

$$(\hat{\pi}_2 | \pi_Y^*)_U = \frac{\hat{\lambda} - \pi_Y^*(1-p)}{p}$$

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

ดังนั้น ค่าความเอนเอียงของ $\text{Bias}(\hat{\pi}_2 | \pi_y^*)_U$ คือ

$$\begin{aligned}\text{Bias}(\hat{\pi}_2 | \pi_y^*)_U &= E\{(\hat{\pi}_2 | \pi_y^*)_U - \pi\} \\ &= \frac{1}{p} \{ [p\pi_{T_A} + (1-p)\pi_y] - (\pi_y - C)(1-p) - p\pi_A \} \\ &= \pi(T_A - 1) + \frac{C(1-p)}{p} \\ &= \beta_A + \beta_y\end{aligned}$$

เมื่อ β_A คือ ค่าเอนเอียงที่เกิดขึ้นเนื่องจาก การไม่ตอบตามความจริงในคำถามที่ปกปิด A

β_y คือ ค่าเอนเอียงจากการประมาณค่า π_y จากการสำรวจที่มีมาก่อนแล้ว

อเบอเนตตีและคณะ (Albermathy et al) [20] ได้นำเทคนิคนี้ไปสำรวจข้อมูลเกี่ยวกับการทำแท้งของประชากรหญิงในเมือง North Carolina โดยต้องการหาสัดส่วนการทำแท้งของหญิงที่อายุ 18-44 ปี ในช่วงปี ค.ศ.1968-1969 (π_1) และหาสัดส่วนการทำแท้งตลอดชีวิตที่ผ่านมาของหญิงที่มีอายุตั้งแต่ 18 ปีขึ้นไป (π_2) ตลอดจนทัศนคติของผู้สำรวจต่อการใช้เครื่องมือนี้ แทนการตอบคำถามโดยตรง จากตัวอย่าง 1251 คน พบว่า โดยรวมสัดส่วนการทำแท้งในช่วงปี ค.ศ.1968-1969 เป็น 0.0342 ซึ่งผลเมืองที่ไม่ใช่ผิวขาวมีสัดส่วนการทำแท้งที่สูงกว่าชนผิวขาว คือ เป็น 0.0681 และ 0.0139 ตามลำดับ และหญิงที่ไม่เคยแต่งงานมีสัดส่วนการทำแท้งที่สูงกว่าหญิงที่แต่งงานแล้ว และยังพบว่าปัจจัยด้านอายุ การศึกษายังผลต่อการทำแท้งอีกด้วย

ผลการสำรวจถึงการการทำแท้งตลอดชีวิตของหญิงที่มีอายุตั้งแต่ 18 ปีขึ้นไป (π_2) พบว่า สัดส่วนการทำแท้งสูงกว่า ในช่วง 1 ปีที่ผ่านมาทุกกรณี ยกเว้นกรณีผู้ที่ไม่เคยแต่งงาน และในผู้ที่อายุน้อยกว่า 31 ปีและผู้ที่ได้รับการศึกษามากกว่า 13 ปี ทั้งนี้อาจเป็นเพราะ ความคลาดเคลื่อนเนื่องจากการสุ่มตัวอย่าง และกลุ่มตัวอย่างไม่เหมือนกัน ในที่นี้ในการหา π_1 ใช้ตัวอย่างเดียว ขณะที่ในการหา π_2 มีการแบ่งตัวอย่างเป็น 2 ตัวอย่างย่อย ทั้งนี้ควรแสดงข้อมูลดิบ เพราะความผิดพลาดอาจเนื่องมาจากจำนวนคนโสดที่ต่างกัน และยังพบว่าคนที่ไม่ใช่ผิวขาวมีสัดส่วนการทำแท้งตลอดชีวิตที่สูงกว่าคนผิวขาวและคนที่อายุ 31 ปีขึ้นไปมีสัดส่วนการทำแท้งที่สูงกว่าคนที่อายุน้อยกว่า 31 ปี นั่นเป็นเพราะอัตราเสี่ยงมากกว่า และคนที่มีการศึกษาน้อยมีสัดส่วนการทำแท้งมาก

ผลการสำรวจความคิดเห็นพบว่า 60 % คิดว่าเครื่องมือจะช่วยปกปิดความลับได้ 17 % ไม่เชื่อในเครื่องมือ 16 % ไม่แน่ใจ นั่นคือผู้ตอบส่วนใหญ่เชื่อมั่นในเครื่องมือการสุ่ม จึงกล้าตอบ

ความจริง และในทางกลับกันหากใช้คำถามโดยตรงเกี่ยวกับคำถามที่มีลักษณะปกปิด เช่น การทำ
 แท้งนี้พบว่า 67 % ผู้ตอบคิดว่าคนทั่วไปจะตอบไม่ตรงตามความจริง 17 % ตอบตามความจริงและ
 20 % ไม่แน่ใจจะไม่ตอบความจริง จากผลการศึกษาชี้ให้เห็นว่าระเบียบวิธีการใช้เทคนิคการตอบ
 สนองการสุ่มเป็นเครื่องมือที่คุ้มค่าและให้ข้อมูลที่น่าเชื่อถือมากกว่าการถามคำถามโดยตรง

นอกจากที่กล่าวมาแล้วข้างต้นแล้ว หลิวและโจว [27] ยังศึกษาพบว่าโดยการให้ผู้ตอบ
 ตอบคำถามหลายครั้งต่อคนในวิธีการของกรีนเบิร์กและคณะ [24] นี้ จะช่วยลดความแปรปรวนของ
 ตัวประมาณ แต่ทั้งนี้อาจจะสร้างความยุ่งยากแต่ในทางปฏิบัติสามารถทำได้ ดังตัวอย่างการสำรวจ
 เกี่ยวกับการทำแท้งที่ Taichun ,Taiwan ที่ทำการถามซ้ำ 3 ครั้งต่อคน สัดส่วนตัวประมาณที่ได้
 มีค่าสูงและความแปรปรวนของตัวประมาณมีค่าต่ำส่งผลให้ตัวประมาณมีประสิทธิภาพมากกว่า
 การถามเพียงครั้งเดียว

2.4.3 สรชัย [15]

การประมาณค่าสัดส่วนของประชากรที่มีลักษณะปกปิดด้วยวิธีการนี้ใช้ขนาด
 ตัวอย่าง และเครื่องมือในการสุ่มเลือกเช่นเดียวกับวิธีการของวอร์เนอร์ในการตอบคำถาม ถ้าผู้
 ถูกสัมภาษณ์สุ่มได้ไฟสีแดงจะต้องตอบคำถามข้อแรกซึ่งเป็นเรื่องปกปิดที่ต้องการศึกษา แต่ถ้าได้
 ไฟสีดำนจะต้องตอบว่า “ใช่” เสมอ ซึ่งเหมือนกับถามคำถามที่สองว่า “ไฟที่ท่านดึงได้เป็นสีดำน
 ใช่หรือไม่” วิธีการนี้ก็คือรูปแบบหนึ่งของการใช้คำถามที่ไม่เกี่ยวข้องกัน เมื่อผู้ถูกสัมภาษณ์ทุกคน
 เข้าใจในวิธีการและตอบคำถามตามความจริงแล้ว การประมาณค่าสัดส่วนประชากรที่มีลักษณะ
 ปกปิดด้วยวิธีนี้น่าจะเป็นสูงสุดหาได้ดังนี้

1) กรณีที่ผู้ถูกสัมภาษณ์ทุกคนตอบความจริง

กำหนดให้ π แทน สัดส่วนที่แท้จริงของประชากรที่มีลักษณะปกปิด

π_s แทน สัดส่วนของตัวอย่างที่มีลักษณะปกปิดโดยวิธีการของสรชัยเมื่อผู้ถูก
 สัมภาษณ์ทุกคนตอบความจริง

p แทน ความน่าจะเป็นที่จะสุ่มได้คำถามปกปิด

$$X = \begin{cases} 1 & \text{ถ้าผู้ถูกสัมภาษณ์ตอบว่า “ใช่”} \\ 0 & \text{ถ้าผู้ถูกสัมภาษณ์ตอบว่า “ไม่ใช่”} \end{cases}$$

ดังนั้น

$$P(X=1) = \pi p + (1-p)$$

$$P(X=0) = 1 - \{\pi p + (1-p)\}$$

ให้ n_1 แทน จำนวนตัวอย่างที่ตอบว่า “ใช่” จากจำนวนผู้ถูกสัมภาษณ์ n คน

$n - n_1$ แทน จำนวนตัวอย่างที่ตอบว่า “ไม่ใช่”

จะได้ ฟังก์ชันภาษาน่าจะเป็น

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

โดยวิธีการประมาณค่าแบบภาษาน่าจะเป็นสูงสุด จะได้ตัวประมาณได้ดังนี้

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

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

$$0 = \frac{n_1 p}{\pi_p + (1-p)} - \frac{(n - n_1)p}{1 - \{\pi_p + (1-p)\}}$$

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

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

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

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

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

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

นั่นคือ ค่าประมาณภาวะน่าจะเป็นสูงสุด ของ π มีค่าเป็นดังนี้

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

ซึ่งเป็นตัวประมาณที่ไม่เอนเอียงดังแสดงต่อไปนี้

$$\begin{aligned} E(\hat{\pi}_s) &= 1 - \frac{1}{p} + \frac{E(n_1)}{pn} \\ &= 1 - \frac{1}{p} + \frac{n[\pi p + (1-p)]}{pn} \\ &= \pi \end{aligned}$$

และความแปรปรวนของ $\hat{\pi}_s$ เป็นดังนี้

$$\begin{aligned} \text{Var}(\hat{\pi}_s) &= \text{Var}\left[1 - \frac{1}{p} + \frac{n_1}{pn}\right] \\ &= \frac{\text{Var}(n_1)}{p^2 n^2} \\ &= \frac{n[\pi p + (1-p)] [1 - \{\pi p + (1-p)\}]}{n^2 p^2} \\ &= \frac{[\pi p + (1-p)] (p - \pi p)}{np^2} \\ &= \frac{[\pi p + (1-p)](1-\pi)}{np} \\ &= \frac{\pi p(1-\pi) + (1-p)(1-\pi)}{np} \end{aligned}$$

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

ค่าประมาณ $\text{Var}(\hat{\pi}_s)$ คือ

$$\hat{\text{Var}}(\hat{\pi}_s) = \frac{\hat{\pi}(1-\hat{\pi})}{n} + \frac{(1-\hat{\pi})(1-p)}{np}$$

นั่นคือ $\hat{\pi}_s$ เป็นตัวประมาณที่ไม่เอนเอียงและ ความแปรปรวนสามารถเขียนในรูปของผลรวมระหว่าง ความแปรปรวนเนื่องมาจากการสุ่มตัวอย่าง และความแปรปรวนเนื่องมาจากเครื่องมือการตอบสนองการสุ่ม ซึ่งจะเห็นว่า $\text{Var}(\hat{\pi}_s)$ หรือ $\hat{\text{Var}}(\hat{\pi}_s)$ จะมีค่าน้อยถ้าจำนวนไฟสีแดง/จำนวนไฟสีดำ มีค่าน้อย นั่นคือถ้าต้องการให้ค่าประมาณสัดส่วนที่หาได้มีความเชื่อถือได้มากที่สุด จะต้องใช้จำนวนไฟสีดำให้น้อยที่สุดเมื่อเทียบกับจำนวนไฟสีแดง ทั้งนี้ไม่ควรให้จำนวนไฟสีดำน้อยเกินไปจนผู้ตอบขาดความไว้วางใจในเครื่องมือ

2) กรณีที่บางส่วนของผู้ถูกสัมภาษณ์ที่มีลักษณะปกปิด ตอบไม่ตรงตามความจริง กำหนดให้ $\hat{\pi}'_s$ แทน สัดส่วนของตัวอย่างที่มีลักษณะปกปิดโดยวิธีสุ่มเมื่อผู้ถูกสัมภาษณ์ที่มีลักษณะปกปิดตอบคำถามไม่ตรงตามความจริง

T_s แทน ความน่าจะเป็นที่ผู้ถูกสัมภาษณ์ที่มีลักษณะปกปิดตอบคำถามปกปิดตามความจริง โดยที่ $0 < T_s \leq 1$

ดังนั้น

$$\begin{aligned} P(X=1) &= \pi p T_s + (1-p) \\ P(X=0) &= 1 - \{\pi p T_s + (1-p)\} \end{aligned}$$

ให้ n_2 แทน จำนวนผู้ถูกสัมภาษณ์ที่ตอบว่า “ใช่” จากจำนวนผู้ถูกสัมภาษณ์ทั้งหมด n คนในทำนองเดียวกันกับ กรณีที่ทุกคนตอบตามความจริง ได้ค่าประมาณ $\hat{\pi}'_s$ คือ

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

เมื่อ

$$\hat{\pi}'_s = \hat{\pi}_s T_s$$

ค่าคาดหวังของ ค่าประมาณ $\hat{\pi}'_s$ คือ

$$\begin{aligned} E(\hat{\pi}'_s) &= 1 - \frac{1}{p} + \frac{E(n_2)}{pn} \\ &= 1 - \frac{1}{p} + \frac{n[\pi p T_s + (1-p)]}{pn} \\ &= 1 - \frac{1}{p} + \frac{[\pi p T_s + (1-p)]}{p} \end{aligned}$$

นั่นคือ $\hat{\pi}'_s$ เป็นตัวประมาณที่เอนเอียงของ π

ดังนั้น ค่าความเอนเอียง และความแปรปรวน $\hat{\pi}'_s$ คือ

$$\begin{aligned} \text{Bias}(\hat{\pi}'_s) &= E(\hat{\pi}'_s) - \pi \\ &= 1 - \frac{1}{p} + \frac{[\pi p T_s + (1-p)]}{p} - \pi \\ &= \pi(T_s - 1) \\ \text{Var}(\hat{\pi}'_s) &= \text{Var}\left[1 - \frac{1}{p} + \frac{n_2}{pn}\right] \\ &= \frac{\text{Var}(n_2)}{p^2 n^2} \\ &= \frac{n[\pi p T_s + (1-p)] [1 - \{\pi p T_s + (1-p)\}]}{n^2 p^2} \\ \text{Var}(\hat{\pi}'_s) &= \frac{\pi T_s (1 - \pi T_s)}{n} + \frac{(1 - \pi T_s)(1-p)}{np} \end{aligned}$$

2.4.4 คลงชาติ [6]

การประมาณค่าสัดส่วนของประชากรที่มีลักษณะปกปิดด้วยวิธีการนี้ใช้ขนาดตัวอย่างและเครื่องมือในการสุ่มเลือกเช่นเดียวกับวิธีการของวอร์เนอร์ในการตอบคำถามนั้น ผู้สัมภาษณ์จะให้ผู้ถูกสัมภาษณ์สับไพ่จนเข้ากันดีแล้วสุ่มไพ่ 2 ครั้ง ครั้งละใบแบบแทนที่ (With replacement) โดยไม่บอกให้ผู้สัมภาษณ์ทราบว่าสุ่มได้ไพ่สีอะไร มีเงื่อนไขในการตอบดังนี้

สุ่มครั้งที่ 1	สุ่มครั้งที่ 2	เงื่อนไขในการตอบ
สีแดง	สีแดง	} ตอบคำถามที่ปกปิดตามความจริง ว่า “ใช่” หรือ “ไม่ใช่”
สีแดง	สีดำ	
สีดำ	สีแดง	→ ตอบ “ใช่”
สีดำ	สีดำ	→ ตอบ “ไม่ใช่”

ซึ่งจากเงื่อนไขในการตอบ เหมือนกับคำถามที่สองที่ว่า “ไฟที่ท่านดึงได้ไบนแรกเป็นสีดำ และไบนที่สองเป็นสีแดง ใช่หรือไม่” โดยกำหนดสัดส่วนของประชากรในคำถามนี้ เท่ากับความน่าจะเป็นที่สุ่มไฟในไบนแรกได้สีดำ และสุ่มไฟในไบนที่สองได้สีแดงกับความน่าจะเป็นที่สุ่มไบนแรกได้สีดำ และไบนที่สองได้สีดำ สำหรับค่าประมาณสัดส่วนประชากรที่มีลักษณะปกปิดด้วยวิธีภาวน่าจะเป็นสูงสุดหาได้ดังนี้

1) กรณีที่ผู้ถูกสัมภาษณ์ทุกคนตอบความจริง

ให้ π	แทน	สัดส่วนที่แท้จริงของประชากรที่มีลักษณะปกปิด
π_m	แทน	สัดส่วนของตัวอย่างที่มีลักษณะปกปิด โดยวิธีของคลซาคีเมื่อผู้ตอบทุกคนตอบตามความจริง
p_r	แทน	ความน่าจะเป็นที่สุ่มได้ไฟไบนแรกได้สีแดงและสุ่มครั้งที่สองได้สีแดงหรือสีดำซึ่งเท่ากับความน่าจะเป็นที่จะสุ่มได้ไฟสีแดง (p) ในวิธีวอร์เนอร์ หรือวิธีสรัชัย หรือ ความน่าจะเป็นที่จะสุ่มได้คำถามปกปิด
p_{br}	แทน	ความน่าจะเป็นที่สุ่มไฟไบนแรกได้ สีดำ และสุ่มไบนที่สองได้ สีแดง
p_{bb}	แทน	ความน่าจะเป็นที่สุ่มไฟไบนแรกได้ สีดำ และสุ่มไบนที่สองได้ สีดำ

$$\text{ดังนั้น } P(X=1) = \pi p + p_{br}$$

$$P(X=0) = 1 - \{\pi p + p_{br}\}$$

ให้ n_1 แทน จำนวนผู้ถูกสัมภาษณ์ที่ตอบว่า “ใช่” จากจำนวนผู้ถูกสัมภาษณ์ n คน
 $n - n_1$ แทน จำนวนผู้ถูกสัมภาษณ์ที่ตอบว่า “ไม่ใช่”

จะได้ ฟังก์ชันภาวน่าจะเป็น

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

โดยวิธีการประมาณค่าแบบภาวน่าจะเป็นสูงสุด จะได้

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

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

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

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

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

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

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

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

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

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

นั่นคือ ค่าประมาณภาวะน่าจะเป็นสูงสุด ของ π เท่ากับ

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

ค่าคาดหวังและความแปรปรวนของตัวประมาณ $\hat{\pi}_m$ เป็นดังนี้

$$\begin{aligned} E(\hat{\pi}_m) &= E\left[\frac{n_1}{np} - \frac{p_{br}}{p}\right] \\ &= \frac{E[n_1]}{np} - \frac{p_{br}}{p} \\ &= \frac{n[\pi p + p_{br}]}{np} - \frac{p_{br}}{p} \\ &= \pi \end{aligned}$$

$$\begin{aligned} \text{Var}(\hat{\pi}_m) &= \text{Var}\left[\frac{n_1}{np} - \frac{p_{br}}{p}\right] \\ &= \frac{\text{Var}(n_1)}{p^2 n^2} \\ &= \frac{n[\pi p + p_{br}][1 - \{\pi p + p_{br}\}]}{p^2 n^2} \end{aligned}$$

นั่นคือ
$$\text{Var}(\hat{\pi}_m) = \frac{\pi(1-\pi)}{n} + \frac{p_{br} - p_{br}^2 + \pi p - \pi p^2 - 2\pi p p_{br}}{p^2 n}$$

จากข้างต้นจะเห็นว่า $\hat{\pi}_m$ เป็นค่าประมาณที่ไม่เอนเอียงของ π และความแปรปรวนของ $\hat{\pi}_m$ สามารถเขียนได้ในรูปของผลรวมระหว่างความแปรปรวนเนื่องจากการสุ่มตัวอย่าง และความแปรปรวนเนื่องจากเครื่องมือการตอบสนองการสุ่ม

2) กรณีที่บางส่วนของผู้ถูกสัมภาษณ์ที่มีลักษณะปกปิด ตอบไม่ตรงตามความจริง กำหนดให้ π'_m แทน สัดส่วนของตัวอย่างที่มีลักษณะปกปิดโดยวิธีการของ คลาสิคกรณีที่บางส่วนของผู้ถูกสัมภาษณ์ที่มีลักษณะปกปิด ตอบไม่ตรงตามความจริง T_m แทน ความน่าจะเป็นที่ผู้ถูกสัมภาษณ์ที่มีลักษณะปกปิดจะตอบคำถามปกปิดตามความจริง โดยที่ $0 < T_m \leq 1$

ดังนั้น

$$P(X=1) = \pi p T_m + p_{br}$$

$$P(X=0) = 1 - \{\pi p T_m + p_{br}\}$$

ให้ n_2 แทน จำนวนผู้ถูกสัมภาษณ์ที่ตอบว่า “ใช่” จากจำนวนผู้ถูกสัมภาษณ์ทั้งหมด n คน ทำนองเดียวกันกับการหาค่าประมาณ π'_m ได้ค่าประมาณของ π คือ

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

เมื่อ

$$\pi'_m = \pi T_m$$

ค่าคาดหวังของ π'_m เป็นดังนี้

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

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

$$= \frac{[\pi p T_m + p_{br}]}{p} - \frac{p_{br}}{p}$$

นั่นคือ π'_m เป็นตัวประมาณที่เอนเอียงของ π

ดังนั้น ค่าความเอนเอียง และความแปรปรวน π'_m คือ

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

$$= \pi T_m - \pi$$

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

$$\begin{aligned}
 \text{Var}(\hat{\pi}'_m) &= \text{Var}\left[\frac{n_2}{np} - \frac{p_{br}}{p}\right] \\
 &= \frac{\text{Var}(n_2)}{p^2 n^2} \\
 &= \frac{n[\pi p T_m + p_{br}][1 - \{\pi p T_m + p_{br}\}]}{n^2 p^2}
 \end{aligned}$$

นั่นคือ
$$\text{Var}(\hat{\pi}'_m) = \frac{\pi T_m (1 - \pi T_m)}{n} + \frac{p_{br} - p_{br}^2 + \pi p T_m - \pi p^2 T_m - 2\pi p p_{br} T_m}{p^2 n}$$

2.4.5 โพลัสัม และคณะ [22]

โพลัสัมและคณะ ได้ประยุกต์เทคนิคการตอบสนองการสุ่มสำหรับข้อมูลเชิงคุณภาพ ให้ π_Y แทนสัดส่วนของประชากรที่มีลักษณะไม่ปกปิด Y ซึ่งไม่ทราบค่าและให้ π_A แทนสัดส่วนของประชากรที่มีลักษณะปกปิด A วิธีการของเทคนิคนี้ใช้ตัวอย่าง 2 กลุ่มและใช้คำถามที่มีลักษณะไม่ปกปิดสองคำถามที่ต่างกัน แทนด้วย Y_1 และ Y_2 แล้วให้แต่ละกลุ่มตอบคำถาม 2 ลักษณะดังนี้

เทคนิคที่ใช้ในการตอบ	ตัวอย่างที่ 1	ตัวอย่างที่ 2
สุ่มคำถามจากเครื่องมือก่อนแล้วตอบคำถามนั้น	คำถาม Sensitive A คำถาม Nonsensitive Y_1	คำถาม Sensitive A คำถาม Nonsensitive Y_2
ตอบคำถามที่ถามโดยตรง	คำถาม Nonsensitive Y_2	คำถาม Nonsensitive Y_1

สุ่มตัวอย่างจากประชากรเดียวกัน 2 กลุ่มแต่ละกลุ่มเป็นอิสระต่อกัน จากนั้นสมาชิกแต่ละคนจากตัวอย่าง 1 ตอบคำถาม Y_2 โดยตรง แล้วสุ่มคำถามจากเครื่องมือที่มีคำถามปกปิด A และคำถาม Y_1 สุ่มแบบใส่กั้นไม่ให้ผู้ใดเห็น หยิบได้คำถามใดให้ตอบคำถามนั้นด้วยความเป็นจริง สำหรับตัวอย่าง 2 ก็เช่นเดียวกันตอบคำถาม Y_1 โดยตรงก่อนแล้วสุ่มคำถามจากเครื่องมือที่มีคำถามปกปิด A และคำถาม Y_2 สุ่มแบบใส่กั้นไม่ให้ผู้ใดเห็น หยิบได้คำถามใดให้ตอบคำถามนั้นด้วยความเป็นจริง ทั้งนี้ทั้งสองตัวอย่างมีโอกาสที่จะเลือกได้คำถามปกปิด A จากเครื่องมือด้วยความน่าจะเป็น p

กำหนดให้ π_A	แทน	สัดส่วนของประชากรที่มีลักษณะปกปิด A
λ_i^r	แทน	ความน่าจะเป็นที่ผู้ตอบจากตัวอย่างที่ i ที่ใช้เครื่องมือการตอบสนองการ สุ่มจะตอบว่า “ใช่” เมื่อ $i = 1, 2$
λ_i^d	แทน	ความน่าจะเป็นที่ผู้ตอบจากตัวอย่างที่ i ตอบคำถามโดยตรงว่า “ใช่” เมื่อ $i = 1, 2$
λ_i^{rd}	แทน	ความน่าจะเป็นที่ผู้ตอบจากตัวอย่างที่ i ตอบว่า “ใช่” เมื่อ $i = 1, 2$
π_{Y_i}	แทน	สัดส่วนของประชากรที่มีลักษณะปกปิด Y_i เมื่อ $i=1,2$
π_{AY_i}	แทน	สัดส่วนของประชากรที่มีลักษณะปกปิด A และไม่ปกปิด Y_i เมื่อ $i=1,2$
$\pi_{Y_1Y_2}$	แทน	สัดส่วนของประชากรที่มีลักษณะไม่ปกปิด Y_1 และ Y_2

ดังนั้น

จากตัวอย่างที่ 1

จากตัวอย่างที่ 2

$$\lambda_1^r = p\pi_A + (1-p)\pi_{Y_1}$$

$$\lambda_2^r = p\pi_A + (1-p)\pi_{Y_2}$$

$$\lambda_1^d = \pi_{Y_1}$$

$$\lambda_2^d = \pi_{Y_2}$$

$$\lambda_1^{rd} = p\pi_{AY_1} + (1-p)\pi_{Y_1Y_2}$$

$$\lambda_2^{rd} = p\pi_{AY_2} + (1-p)\pi_{Y_1Y_2}$$

ดังนั้น

$$\hat{\lambda}_1^r = p\hat{\pi}_{A_1} + (1-p)\hat{\pi}_{Y_1} \quad (2.13)$$

$$\hat{\lambda}_1^d = \hat{\pi}_{Y_1} \quad (2.14)$$

$$\hat{\lambda}_2^r = p\hat{\pi}_{A_2} + (1-p)\hat{\pi}_{Y_2} \quad (2.15)$$

$$\hat{\lambda}_2^d = \hat{\pi}_{Y_2} \quad (2.16)$$

ต้องการหา $\hat{\pi}_{A_1}$; จาก (2.16) แทนค่า $\hat{\pi}_{Y_1} = \hat{\lambda}_1^d$ ใน (2.13) จะได้

$$\hat{\pi}_{A_1} = \frac{\hat{\lambda}_1^r - (1-p)\hat{\lambda}_1^d}{p}$$

ต้องการหา $\hat{\pi}_{A_2}$; จาก (2.14) แทนค่า $\hat{\pi}_{Y_2} = \hat{\lambda}_2^d$ ใน (2.15) จะได้

$$\hat{\pi}_{A_2} = \frac{\hat{\lambda}_2^r - (1-p)\hat{\lambda}_2^d}{p}$$

ซึ่งจะได้ตัวประมาณพารามิเตอร์สองค่าจากตัวอย่างที่ 1 และ 2 คือ $\hat{\pi}_{A_1}$ และ $\hat{\pi}_{A_2}$ ตามลำดับ กำหนดให้ $(\hat{\pi}_A)_{u2}$ คือตัวประมาณพารามิเตอร์ π_A ที่เหมาะสมซึ่ง

$$(\hat{\pi}_A)_{u2} = w\hat{\pi}_{A_1} + (1-w)\hat{\pi}_{A_2}$$

และ $(\hat{\pi}_A)_{u2}$ เป็นตัวประมาณที่ไม่เอนเอียง

ค่า w ที่ทำให้ความแปรปรวนของตัวประมาณมีค่าน้อยที่สุด (w_{opt}) คือ

$$w_{opt} = (\sum_2^2 - \sum_{12}) / \left(\sum_1^2 + \sum_2^2 - 2\sum_{12} \right)$$

เมื่อ $\sum_1^2 = \text{Var}\{\hat{\pi}_A(1)\} = (1/p^2) \{ \lambda_1^r (1 - \lambda_1^r) / n_1 + (1-p)^2 \pi_{Y_1} (1 - \pi_{Y_1}) / n_2 \}$

$$\sum_2^2 = \text{Var}\{\hat{\pi}_A(2)\} = (1/p^2) \{ \lambda_2^r (1 - \lambda_2^r) / n_2 + (1-p)^2 \pi_{Y_1} (1 - \pi_{Y_1}) / n_1 \}$$

$$\begin{aligned} \sum_{12} &= \text{COV}\{\hat{\pi}_A(1); \hat{\pi}_A(2)\} \\ &= \frac{-(1-p)}{p^2} \{ (\lambda_1^r - \lambda_1^r \pi_{Y_1}) / n_1 + (\lambda_2^r - \lambda_2^r \pi_{Y_1}) / n_2 \} \end{aligned}$$

ซึ่งได้ค่าความแปรปรวนที่น้อยที่สุดเป็นดังนี้

$$\text{Minimum Var}(\hat{\pi}_A)_{u2} = \left(\sum_1^2 \sum_2^2 - \sum_{12}^2 \right) / \left(\sum_1^2 + \sum_2^2 - 2\sum_{12} \right)$$

ต่อมา ชิมิซุและบอนแฮม (Shimizu and Bonham) [30] ได้นำเทคนิคนี้มาใช้ในการเก็บข้อมูลเพื่อใช้ในการหาสัดส่วนของผู้หญิงที่ทำแท้งในช่วงเวลา 12 เดือนปี ค.ศ.1973 ในสหรัฐอเมริกา โดยประชากรเป็นผู้หญิงที่มีอายุ 15-44 ปีอาศัยอยู่ในเขตเมืองซึ่งเป็นหญิงที่แต่งงานแล้วหรือมีลูกแล้ว โดยแบ่งตัวอย่างออกเป็นสองกลุ่มคือตัวอย่างที่ 1 เป็นกลุ่มที่มีเลขสุดท้ายของเลขที่บ้านเป็นเลขคี่และตัวอย่าง 2 คือ กลุ่มที่มีเลขสุดท้ายของเลขที่บ้านเป็นเลขคู่ ในการเลือกคำถามใน

เครื่องมือที่ใช้หรือยืมเป็นเครื่องมือในการสุ่มนั้นคือ $p=0.5$ ซึ่งจากงานวิจัยนี้ใช้คำถามดังตารางต่อไป

ผลการโยนเหรียญ	ตัวอย่างที่1	ตัวอย่างที่2
ขึ้นหัว	Q_A : ช่วง 12 เดือนที่ผ่านมา คุณเคยทำแท้งใช่หรือไม่	Q_A : ช่วง 12 เดือนที่ผ่านมา คุณเคยทำแท้งใช่หรือไม่
ขึ้นก้อย	Q_{Y1} : คุณเกิดเดือนเมษายน ใช่หรือไม่	Q_{Y2} : ปีที่ผ่านมาคุณเคยย้ายถิ่น ใช่หรือไม่
คำถามโดยตรง	Q_{Y2} : ปีที่ผ่านมาคุณเคยย้ายถิ่น ใช่หรือไม่	Q_{Y1} : คุณเกิดเดือนเมษายน ใช่หรือไม่

ผลจากเทคนิคนี้ให้ตัวเลขของสัดส่วนของผู้ที่เกยท่าแท้งที่สูงกว่าวิธีการถามโดยตรง และข้อมูลน่าเชื่อถือและครอบคลุมกว่าที่เก็บไว้ในกองทะเบียน ทั้งนี้เพราะข้อมูลที่ได้จากการเก็บของกองทะเบียนอาจมีการบกพร่องในการรายงานเพราะการทำแท้งยังเป็นเรื่องที่ผิดกฎหมายอยู่ ผู้ทำอาจมีความผิดเชิงอาญาได้ นอกจากว่าการทำแท้งนั้นได้รับความเห็นชอบจากแพทย์ แต่ทั้งนี้ในการเก็บข้อมูลโดยใช้เครื่องมือการตอบสนองการสุ่ม ยังพบข้อผิดพลาดที่อาจจะเกิดได้จากการตอบคำถามทั้งคำถามปิดหรือ คำถามที่ไม่เกี่ยวข้องที่ผู้ตอบอาจไม่เข้าใจ หรือไม่ทราบคำตอบที่แท้จริงซึ่งเป็นเรื่องที่ต้องปรับปรุงเครื่องมือต่อไป

2.5 เทคนิคการตอบสนองการสุ่ม สำหรับข้อมูลเชิงปริมาณ

ในการเก็บรวบรวมข้อมูลบางครั้งเป็นข้อมูลเชิงปริมาณเช่น รายได้ของสมาชิกในครอบครัว จำนวนครั้งในการทำแท้ง จำนวนคนที่มิเพศสัมพันธ์กันก่อนแต่งงานเป็นต้น จึงได้มีผู้คิดวิธีการเก็บข้อมูลที่มีลักษณะดังกล่าวหลายคณะคือ

2.5.1 กรีนเบิร์กและคณะ

2.5.2 หลิวและคณะ

2.5.3 หลิวและโจว

ดังแสดงรายละเอียดต่อไปนี้

2.5.1 กรีนเบิร์กและคณะ [24]

อาบูอีลาและคณะ (Abul-Ela et al) [19] ได้พัฒนาแนวคิดของวอร์เนอร์สำหรับ trichotomous case ซึ่งเป็นการประมาณสัดส่วนของ 3 เรื่องที่ปกปิดที่เกี่ยวข้องกันแต่ไม่เกิดร่วมกัน (Mutually Exclusive Groups) และขยายไปสู่การประมาณสัดส่วนของ j กลุ่ม เมื่อ $j \geq 3$ อีกด้วย แต่เป็นเพียงแนวคิดไม่ได้กล่าวถึงการปฏิบัติจริง ต่อมากรีนเบิร์กและคณะ [24] ได้ประยุกต์แนวคิดนี้จากการเก็บข้อมูลเชิงคุณภาพ เป็นการเก็บข้อมูลเชิงปริมาณ ซึ่งเครื่องมือคือกล่องโปรงใสที่บรรจุบอลสีแดงและสีน้ำเงินในอัตราส่วนที่ไม่เท่ากัน ถ้าผู้ตอบสุ่มลูกบอลได้สีแดงให้ตอบคำถามที่มีลักษณะปกปิด (Q_A) ถ้าสุ่มได้ลูกบอลสีน้ำเงินให้ตอบคำถามที่ไม่ปกปิด (Q_Y) ลักษณะคำถาม เช่น

Q_A : “ท่านเคยทำแท้งมาแล้วทั้งหมดกี่ครั้ง”

Q_Y : “ท่านคิดว่าคนที่มียานประจำควรมีบุตรทั้งหมดกี่คน”

คำตอบที่ได้จะเป็นข้อมูลเชิงปริมาณ โดยในการตอบ ผู้ถามจะไม่ทราบว่าผู้ตอบสุ่มได้บอลสีอะไร การดำเนินการจะใช้ตัวอย่าง 2 กลุ่มขนาด n_1 และ n_2 ตามลำดับ ที่เป็นอิสระต่อกันและไม่เกี่ยวข้องกัน (Two independent, non-overlapping samples) แล้วนำคำตอบที่ได้ไปประมาณค่าเฉลี่ย ที่แท้จริง (μ_A) ของประชากรที่มีลักษณะปกปิดโดยใช้หลักของสถิติและความน่าจะเป็น ซึ่งมีข้อสมมติเบื้องต้นว่าทุกคนตอบคำถามตามความจริง ดังรายละเอียดการประมาณค่าต่อไปนี้

กำหนดให้

p_i แทน ความน่าจะเป็นที่ตัวอย่าง i สุ่มได้คำถามที่ปกปิด เมื่อ $i=1,2$

Z_i แทน คำตอบจากตัวอย่างที่ i ($i=1,2$)

$f(z_i)$ แทน ฟังก์ชันความน่าจะเป็นของคำตอบจากคำถามปกปิด ; $E_f(Z_i) = \mu_A$

$g(z_i)$ แทน ฟังก์ชันความน่าจะเป็นของคำตอบจากคำถามไม่ปกปิด ; $E_g(Z) = \mu_Y$

μ_A แทน ค่าเฉลี่ยคำตอบของประชากรที่มีลักษณะปกปิด

μ_Y แทน ค่าเฉลี่ยของคำตอบของประชากรที่มีลักษณะไม่ปกปิด

$\hat{\mu}_A$ แทน ค่าเฉลี่ยของคำตอบตัวอย่างที่มีลักษณะปกปิด

$\hat{\mu}_Y$ แทน ค่าเฉลี่ยของคำตอบตัวอย่างที่มีลักษณะไม่ปกปิด

ฟังก์ชันความน่าจะเป็นของคำตอบแต่ละตัวอย่างคือ

$$\text{ตัวอย่าง 1: } \psi_1(z_1) = p_1 f(z_1) + (1 - p_1) g(z_1)$$

$$\text{ตัวอย่าง 2: } \psi_2(z_2) = p_2 f(z_2) + (1 - p_2) g(z_2)$$

โดยที่

$$\begin{aligned}\mu_{z_1} &= E(Z_1) \\ &= p_1\mu_A + (1-p_1)\mu_Y\end{aligned}\quad (2.17)$$

$$\begin{aligned}\mu_{z_2} &= E(Z_2) \\ &= p_2\mu_A + (1-p_2)\mu_Y.\end{aligned}\quad (2.18)$$

แก้สมการ(2.17) และ (2.18) จะได้ μ_A และ μ_Y ดังนี้

$$\begin{aligned}\mu_A &= \frac{(1-p_2)\mu_{z_1} - (1-p_1)\mu_{z_2}}{p_1 - p_2} \\ \mu_Y &= \frac{p_2\mu_{z_1} - p_1\mu_{z_2}}{p_2 - p_1}\end{aligned}$$

แทนค่า μ_{z_1}, μ_{z_2} ด้วย $\bar{Z}_1, \bar{Z}_2$ ตามลำดับเพื่อประมาณ μ_A, μ_Y ได้ดังนี้

$$\begin{aligned}\hat{\mu}_A &= \frac{(1-p_2)\bar{Z}_1 - (1-p_1)\bar{Z}_2}{p_1 - p_2} \\ \hat{\mu}_Y &= \frac{p_2\bar{Z}_1 - p_1\bar{Z}_2}{p_2 - p_1}\end{aligned}$$

ซึ่ง $\hat{\mu}_A, \hat{\mu}_Y$ เป็นตัวประมาณที่ไม่เอนเอียงของ μ_A, μ_Y

และความแปรปรวนของตัวประมาณ μ_A, μ_Y คือ

$$\begin{aligned}\text{Var}(\hat{\mu}_A) &= \frac{1}{(p_1 - p_2)^2} \left[(1-p_2)^2 \text{Var}(\bar{Z}_1) + (1-p_1)^2 \text{Var}(\bar{Z}_2) \right] \\ \text{Var}(\hat{\mu}_Y) &= \frac{1}{(p_2 - p_1)^2} \left[p_2^2 \text{Var}(\bar{Z}_1) + p_1^2 \text{Var}(\bar{Z}_2) \right]\end{aligned}$$

เมื่อ

$$\text{Var}(\bar{Z}_i) = \frac{1}{n_i} \left[E(Z_i^2) - \{E(Z_i)\}^2 \right]$$

$$\begin{aligned}
&= \frac{1}{n_i} \left[\{p_i E_f(Z_i^2) + (1-p_i) E_g(Z_i^2)\} - \{p_i \mu_A + (1-p_i) \mu_Y\}^2 \right] \\
&= \frac{1}{n_i} \left[p_i \sigma_A^2 + (1-p_i) \sigma_Y^2 + p_i \mu_A^2 + (1-p_i) \mu_Y^2 - \{p_i \mu_A + (1-p_i) \mu_Y\}^2 \right] \\
\text{Var}(\bar{Z}_i) &= \frac{1}{n_i} \left[\sigma_Y^2 + p_i (\sigma_A^2 - \sigma_Y^2) + p_i (1-p_i) (\mu_A - \mu_Y)^2 \right], i=1,2
\end{aligned}$$

ในกรณีที่ทราบค่า μ_Y, σ_Y สามารถหาตัวประมาณจาก

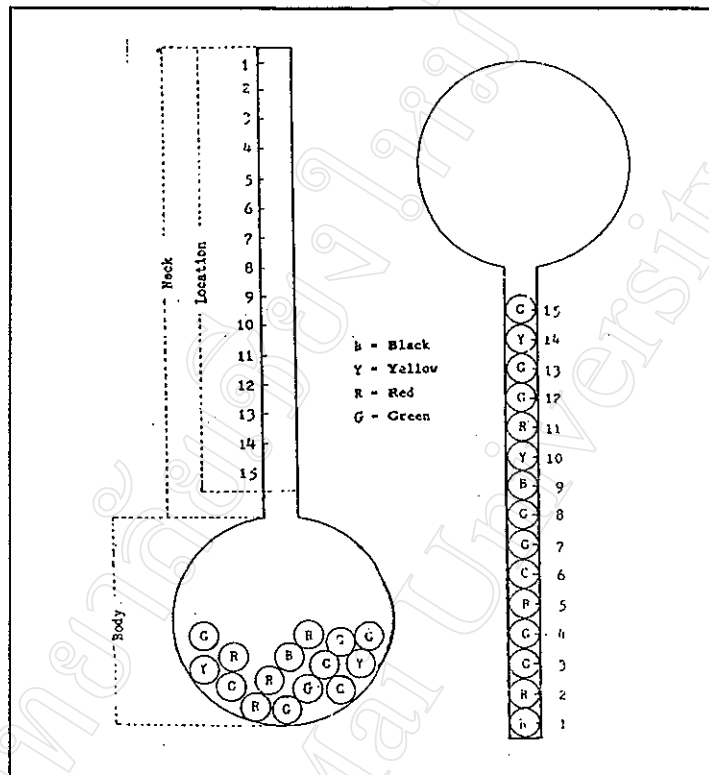
$$\begin{aligned}
\hat{\mu}_A | \mu_Y &= \frac{[\bar{Z} - (1-p) \mu_Y]}{p} \\
\text{Var}(\hat{\mu}_A | \mu_Y) &= \frac{\text{Var}(\bar{Z})}{p^2} \\
&= \frac{\text{Var}(Z)}{np^2}
\end{aligned}$$

จากการเก็บข้อมูลการทำแท้งของประชากรหญิงในเมือง North Carolina ตลอดปี ค.ศ.1968 พบว่าหญิงชนผิวขาวมีการทำแท้งโดยเฉลี่ย 0.415 ที่ความเบี่ยงเบนมาตรฐาน 0.107 ในขณะที่หญิงที่ไม่ใช่ชนผิวขาวมีการทำแท้งที่สูงกว่า คือมีการทำแท้งเฉลี่ย 0.645 ที่ความเบี่ยงเบนมาตรฐาน 0.117 และมีการเก็บข้อมูลเกี่ยวกับรายได้เฉลี่ยของหัวหน้าครอบครัวได้ในการสำรวจที่ในเมือง North Carolina ตลอดปี ค.ศ. 1968 พบว่าคนผิวขาวมีรายได้เฉลี่ย 6,690 ดอลลาร์และที่ไม่ใช่คนผิวขาว มีรายได้เฉลี่ย 3,370 ดอลลาร์ โดยส่วนเบี่ยงเบนมาตรฐานเป็น 418 และ 445 ดอลลาร์ ตามลำดับ

2.5.2 หลิวและคณะ [28]

หลิวและคณะได้ศึกษาเทคนิคการตอบสนองการสุ่ม เมื่อสิ่งที่สนใจเป็นข้อมูลเชิงปริมาณแบบไม่ต่อเนื่อง โดยให้หน่วยตัวอย่างใช้เครื่องมือที่เรียกว่า Hopkins Randomizing Device- Model 3 ซึ่งมีลักษณะเป็นขวดแก้วคอยาวที่บรรจุลูกบอล k สี แต่ละสีจำนวนไม่เท่ากัน โดยให้ $m_1, m_2, m_3, \dots, m_k$ แทนจำนวนสีที่ 1, 2, 3, ..., k ตามลำดับและ $m_1 \neq m_2 \neq m_3 \dots \neq m_k$ โดยที่แต่ละสีจะแทนลักษณะปกติของกลุ่มนั้น นั่นคือ ให้สีที่ 1 แทนลักษณะของกลุ่มที่ 1 สีที่ 2 แทนลักษณะของกลุ่มที่ 2 ... และสีที่ k แทนลักษณะของกลุ่มที่ k

ลักษณะเครื่องมือ (Hopkins Randomizing Device- Model 3) เป็นดังรูป



รูป 2.1 Hopkins Randomizing Device- Model 3

การดำเนินการให้ผู้ตอบเขย่าขวดเพื่อให้ลูกบอลแต่ละลูกมีโอกาสตกลงในคอขวดเท่าๆ กัน แล้วคว่ำขวดแล้วให้ลูกบอลตกในคอขวดที่มีอยู่ m ตำแหน่ง แล้วให้ผู้ถูกสัมภาษณ์ตอบถึง ตำแหน่งของสีบอลในสีที่ผู้ตอบอยู่ ปรากฏเป็นตำแหน่งแรกในคอขวด เช่น ต้องการศึกษาว่า นักศึกษาเคยมีคูรักรมาแล้วกี่คน กำหนดให้สีดำแทนไม่เคยมีคูรักรมาก่อนเลย สีเหลืองแทนเคยมี คูรักร 1 คน สีแดงแทนเคยมีคูรักร 2 คน สีเขียวแทนเคยมีคูรักรตั้งแต่ 3 คนขึ้นไป ในเครื่องมือให้มี บอลอยู่ 15 ลูก ประกอบด้วย สีดำ สีเหลือง สีแดง สีเขียว จำนวน 1, 2, 4 และ 8 ตาม ลำดับ ถ้าผู้ถูก สัมภาษณ์ ไม่เคยมีคูรักรมาก่อนเลย ให้ตอบตำแหน่งของสีดำที่ปรากฏเป็นตำแหน่งแรกในคอขวด ในที่นี้ถ้าเขย่าแล้วดูตำแหน่งในรูปก็ตอบ 2 แต่ถ้าผู้ถูกสัมภาษณ์ เคยมีคูรักร 1 คน ให้ตอบตำแหน่ง ของสีเหลืองที่ปรากฏเป็นตำแหน่งแรกในคอขวดซึ่งต้องตอบ 10 เป็นต้น เมื่อผู้ถูกสัมภาษณ์ ทุกคนเข้าใจแล้วใช้เครื่องมือแล้วตอบตามความจริง จากนั้นนำหลักสถิติและความน่าจะเป็น มาประมาณค่าพารามิเตอร์ดังรายละเอียดต่อไปนี้

กำหนดให้

m_i แทน จำนวนบอลสีที่ i

m แทน จำนวนบอลในขวดทั้งหมด ซึ่งเท่ากับ $\sum_{i=1}^k m_i$

$X_{i,j}$ แทน ความน่าจะเป็นที่สี i ปรากฏเป็นครั้งแรกในตำแหน่งที่ j

เช่น $X_{i,1}$ คือความน่าจะเป็นที่สี i ปรากฏเป็นครั้งแรกในตำแหน่งที่ 1

$$X_{i,1} = \frac{m_i}{m}$$

$X_{i,2}$ คือความน่าจะเป็นที่สี i ปรากฏเป็นครั้งแรกในตำแหน่งที่ 2 ซึ่งเท่ากับ ความน่าจะเป็นที่สี i ปรากฏในตำแหน่งที่ 2 เมื่อ สี i ไม่ปรากฏในตำแหน่งที่ 1

$$X_{i,2} = \left(\frac{m - m_i}{m} \right) \left(\frac{m_i}{m - 1} \right)$$

$X_{i,3}$ คือความน่าจะเป็นที่สี i ปรากฏเป็นครั้งแรกในตำแหน่งที่ 3 ซึ่งเท่ากับ ความน่าจะเป็นที่สี i ปรากฏในตำแหน่งที่ 3 เมื่อ สี i ไม่ปรากฏในตำแหน่งที่ 1 และ 2 มาก่อน

$$X_{i,3} = \left(\frac{m - m_i}{m} \right) \left(\frac{m - m_i - 1}{m - 1} \right) \left(\frac{m_i}{m - 2} \right)$$

ในทำนองเดียวกัน $X_{i,j}$ แทนความน่าจะเป็นที่สี i ปรากฏเป็นครั้งแรกในตำแหน่งที่ j คือ

$$X_{i,j} = \begin{cases} \frac{((m-j)!(m-m_i)m_i)}{(m!(m-m_i+1-j))}; j=1,2,3,\dots,(m-m_i+1) \\ 0 \quad \text{เมื่อเป็นค่าอื่น} \end{cases}$$

π_i แทนสัดส่วนที่แท้จริงของกลุ่ม i ในประชากร ($\sum_{i=1}^k \pi_i = 1$)

P_j แทนความน่าจะเป็นที่ผู้ตอบจะตอบว่า j

ดังนั้น $P_j = \sum_{i=1}^k \pi_i X_{i,j}$ เมื่อ $j=1,2,3,\dots,(m-m_i+1)$

ถ้าสุ่มตัวอย่างขนาด n และให้ n_j คือจำนวนผู้ที่ตอบว่า j ให้ Y_j แทนสัดส่วนของผู้ที่ตอบว่า j

ดังนั้น
$$Y_j = \frac{n_j}{n}$$

เมื่อ
$$E(Y_j) = P_j = \sum_{i=1}^k \pi_i X_{i,j}$$

ดังนั้นเขียนในรูปสมการเชิงเส้นอย่างง่ายได้

$$\begin{aligned} Y_j &= \sum_{i=1}^k \pi_i X_{i,j} + e_j \\ &= \sum_{i=1}^k \pi_i X_{i,j} + \pi_k X_{k,j} + e_j \\ &= \sum_{i=1}^k \pi_i X_{i,j} + \left(1 - \sum_{i=1}^{k-1} \pi_i\right) X_{k,j} + e_j \\ &= \sum_{i=1}^k \pi_i X_{i,j} + X_{k,j} - X_{k,j} \sum_{i=1}^{k-1} \pi_i + e_j \end{aligned}$$

ดังนั้น
$$Y_j - X_{k,j} = \sum_{i=1}^k \pi_i (X_{i,j} - X_{k,j}) + e_j$$

เขียนในรูป สมการเมทริกซ์พีชคณิตได้คือ

$$Y = X\pi + e$$

เมื่อ
$$Y = [Y_j - X_{k,j}]$$
 เป็นเมทริกซ์ขนาด $m \times 1$

$$X = [X_{i,j} - X_{k,j}]$$
 เป็นเมทริกซ์ขนาด $m \times (k-1)$

$$\pi = [\pi_i]$$
 เป็นเมทริกซ์ขนาด $(k-1) \times 1$

$$e = [e_j]$$
 เป็นเมทริกซ์ขนาด $m \times 1$

เนื่องจาก $E(e)=0$ และเมทริกซ์ของความแปรปรวนและความแปรปรวนร่วมของ e (variance – covariance matrix) คือ $V=V_{i,j}$ ซึ่งเป็นเมทริกซ์ขนาด $m \times m$ โดยที่

$$V_{i,i} = \frac{P_i(1-P_i)}{n} \text{ และ } V_{i,j} = \frac{-P_i P_j}{n}$$

สามารถประมาณค่าพารามิเตอร์โดยใช้วิธีการ Weighted Least Squares ได้ดังนี้

$$\hat{\pi} = (X'V^{-1}X)^{-1}X'V^{-1}Y \text{ และ } \hat{\pi}_k = 1 - \sum_{i=1}^{k-1} \hat{\pi}_i$$

สามารถหาความแปรปรวนได้ดังนี้

$$\text{Var}(\hat{\pi}) = (X'V^{-1}X)^{-1} = C \text{ เมื่อ } C \text{ เป็นเมทริกซ์ขนาด } m \times 1$$

$$\text{และ } \text{Var}(\hat{\pi}_k) = 1 - \sum_{i=1}^{k-1} C_{i,i} + 2 \sum_{i \neq j}^{k-1} C_{i,j}$$

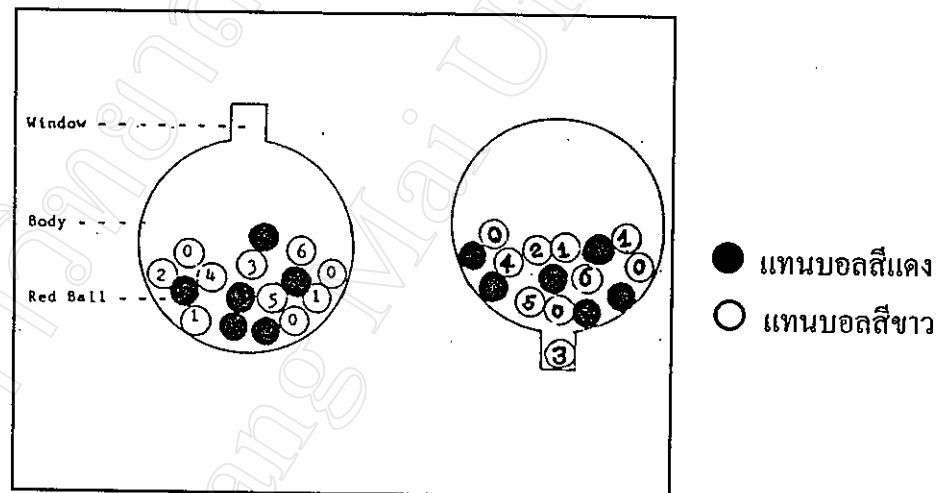
ตัวอย่างการประยุกต์ใน [28] โดยการสำรวจนักศึกษาใน Jonhs Hopkins University School of Hygiene and Public Health จำนวน 34 คนทั้งชายและหญิง ถามถึงจำนวนรุ่นนอนที่เคยมีเพศสัมพันธ์ด้วยตลอดชีวิตที่ผ่านมาจนกระทั่งแต่งงานด้วย โดยในเครื่องมือมีบอล 15 ลูกในเครื่องมือให้มือบอลอยู่ 15 ลูก ประกอบด้วย สีดำ สีเหลือง สีแดง สีเขียว จำนวน 1, 2, 4 และ 8 ตามลำดับ ให้สีดำ แทนไม่เคยมีรุ่นนอนมาก่อนเลย สีเหลืองแทนเคยมีรุ่นนอน 1 คน สีแดงแทนเคยมีรุ่นนอน 2 คน สีเขียวแทนเคยมีรุ่นนอนตั้งแต่ 3 คนขึ้นไป ซึ่งผลการสำรวจพบว่านักศึกษาไม่เคยมีรุ่นนอนมาก่อนเลย 9.3 % เคยมีรุ่นนอน 1 คน 42.2 % เคยมีรุ่นนอน 2 คนคิดเป็น 18 % และเคยมีรุ่นนอนตั้งแต่ 3 คนขึ้นไป 30.5%

2.5.3 หลิวและโจว [26]

หลิวและโจวได้ศึกษาเทคนิคการตอบสนองการสุ่มเมื่อสิ่งที่สนใจเป็นข้อมูลเชิงปริมาณ แบบไม่ต่อเนื่อง โดยให้หน่วยตัวอย่างใช้เครื่องมือที่เรียกว่า Hopkins Randomizing

Device- Model 4 ซึ่งมีลักษณะเป็นขวดแก้วคอสั้นที่บรรจุลูกบอล 2 สีที่แตกต่างกันและแต่ละสีจำนวนไม่เท่ากัน โดยที่สีหนึ่งแทนคำถามที่มีลักษณะปกปิดกลุ่มที่ 1 อีกสีหนึ่งเขียนหมายเลข $0, 1, 2, 3, \dots, k$ และแต่ละหมายเลขมีจำนวนลูกบอลไม่เท่ากัน ตัวอย่างเครื่องมือคือภายในขวดมีบอลสีแดง และขาวในอัตราส่วนที่ต่างกันบนบอลสีขาวจะมีหมายเลขปรากฏ วิธีการใช้เครื่องมือคือให้หน่วยตัวอย่างเขย่าขวดให้ลูกบอลที่อยู่ในขวดคลุกเคล้ากันจากนั้นคว้าขวดลงโดยให้มีลูกบอลเพียงหนึ่งลูกตกลงมาในคอขวด ถ้าลูกบอลที่ตกลงมาเป็นสีแดงให้ผู้ตอบคำถามปกปิด เช่น จำนวนครั้งที่เคยทำแท้ง เป็นต้น ถ้าผู้ตอบได้ลูกบอลสีขาวให้ตอบหมายเลขที่ปรากฏบนลูกบอล เช่น ถ้าเป็นบอลสีขาวเลขหนึ่งก็ตอบ “หนึ่ง” ทั้งนี้ในการตอบคำถามผู้ถามจะไม่ทราบว่าเป็นบอลที่ตกลงมาเป็นสีอะไรใช้เครื่องมือดังกล่าวแล้วนำผลทั้งหมดไปประมาณค่าสัดส่วน (π) ที่แท้จริง

ลักษณะเครื่องมือเป็นดังรูป



รูป 2.2 Hopkins Randomizing Device- Model 4

ให้ w_i แทน จำนวนบอลสีขาวของกลุ่ม i เมื่อ $i = 0, 1, 2, 3, 4, \dots, k$

r แทน จำนวนบอลสีแดง

จำนวนลูกบอลทั้งหมดในขวดคือ $r + \sum_{i=0}^k w_i = r + w; \sum_{i=0}^k w_i \neq 0$

π_i แทน สัดส่วนที่แท้จริงของผู้ที่ตอบ i เช่น π_1 คือสัดส่วนที่แท้จริงของผู้ที่ทำแท้งหนึ่งครั้ง

P_i แทน ความน่าจะเป็นที่จะมีผู้ตอบว่า i

ดังนั้นความน่าจะเป็นของผู้ที่ตอบว่า i คือ

$$P_i = \pi_i \left(\frac{r}{r+w} \right) + \frac{w_i}{r+w} \quad (2.19)$$

ถ้าผู้ตอบมีขนาด n และ n_i แทนจำนวนผู้ที่ตอบว่า i

Y_i แทนสัดส่วนของผู้ที่ตอบว่า i และ Y_i เป็นตัวประมาณของ P_i และนำไปใช้ประมาณ π_i ต่อไป

ดังนั้น สัดส่วนของผู้ที่ตอบว่า $i = \frac{\text{จำนวนผู้ที่ตอบว่า } i}{\text{จำนวนผู้ตอบทั้งหมด}}$

$$\text{นั่นคือ} \quad Y_i = \frac{n_i}{n}$$

จาก (2.19) แทนค่า P_i ด้วย Y_i

$$Y_i = \hat{\pi}_i \left(\frac{r}{r+w} \right) + \frac{w_i}{r+w}$$

$$\hat{\pi}_i = \left(\frac{r+w}{r} \right) Y_i - \frac{w_i}{r}$$

ซึ่งเป็นตัวประมาณที่ไม่เอนเอียงของ π_i ซึ่งแสดงได้ดังนี้

$$\begin{aligned} E(\hat{\pi}_i) &= E \left[\left(\frac{r+w}{r} \right) Y_i - \frac{w_i}{r} \right] \\ &= \left(\frac{r+w}{r} \right) E(Y_i) - \frac{w_i}{r} \end{aligned}$$

พิจารณา

$$\begin{aligned} E(Y_i) &= E \left(\frac{n_i}{n} \right) \\ &= \frac{1}{n} E(n_i) \\ &= \frac{1}{n} \left[n \left(\pi_i \left(\frac{r}{r+w} \right) + \frac{w_i}{r+w} \right) \right] \end{aligned}$$

ดังนั้น

$$\begin{aligned} E(\hat{\pi}_i) &= \frac{r+w}{r} \left[\pi_i \left(\frac{r}{r+w} \right) + \frac{w_i}{r+w} \right] - \frac{w_i}{r} \\ &= \pi_i + \frac{w_i}{r} - \frac{w_i}{r} \\ &= \pi_i \end{aligned}$$

และสามารถหาความแปรปรวนของ $\hat{\pi}_i$ ได้ดังนี้

$$\begin{aligned}\text{Var}(\hat{\pi}_i) &= \text{Var}\left[\left(\frac{r+w}{r}\right)Y_i - \frac{w_i}{r}\right] \\ &= \left(\frac{r+w}{r}\right)^2 \text{Var}(Y_i)\end{aligned}\quad (2.20)$$

จาก (2.20) พิจารณา

$$\begin{aligned}\text{Var}(Y_i) &= \text{Var}\left(\frac{n_i}{n}\right) \\ &= \frac{1}{n^2} \text{Var}(n_i) \\ &= \frac{1}{n^2} \left[n \left(\pi_i \left(\frac{r}{r+w} \right) + \frac{w_i}{r+w} \right) \left(1 - \pi_i \left(\frac{r}{r+w} \right) - \frac{w_i}{r+w} \right) \right] \\ &= \frac{1}{n(r+w)} (\pi_i r + w_i) (r+w - r\pi_i - w_i)\end{aligned}\quad (2.21)$$

แทน (2.21) ใน (2.20) ดังนั้น

$$\text{Var}(\hat{\pi}_i) = \left(\frac{r+w}{r}\right)^2 \left\{ \frac{1}{n(r+w)} (\pi_i r + w_i) (r+w - \pi_i r - w_i) \right\}$$

$$\text{Var}(\hat{\pi}_i) = \frac{r+w}{n} (\pi_i r + w_i) (r+w - \pi_i r - w_i)$$

และสามารถประมาณค่าความแปรปรวนของตัวสถิติ $\hat{\pi}_i$ (แทนด้วย $\hat{\text{Var}}(\hat{\pi}_i)$) ได้ดังนี้

จากสมการ (2.20)

$$\begin{aligned}\hat{\text{Var}}(Y_i) &= \hat{\text{Var}}\left(\frac{n_i}{n}\right) \\ &= \frac{1}{n^2} \hat{\text{Var}}(n_i) \\ &= \frac{1}{n^2} [nY_i(1-Y_i)]\end{aligned}$$

$$= \frac{1}{n} Y_i (1 - Y_i)$$

ดังนั้น
$$\hat{\text{Var}}(\hat{\pi}_i) = \left(\frac{r+w}{r} \right)^2 \left(\frac{1}{n} Y_i (1 - Y_i) \right) \quad (2.22)$$

จากสมการ (2.22) จะเห็นว่าอัตราส่วนของบอลลีแดงต่อจำนวนบอลลีทั้งหมดในเครื่องมือเป็นส่วนที่ควบคุมความแปรปรวน ซึ่งมีผลต่อประสิทธิภาพในการประมาณค่า นั่นคือ ถ้าให้จำนวนบอลลีทั้งหมดคงที่ แต่เพิ่มจำนวนบอลลีแดงจากเดิม ทำให้การประมาณค่ามีประสิทธิภาพมากขึ้น หรือเพิ่มจำนวนบอลลีทั้งหมดแต่ อัตราส่วนบอลลีทั้งหมดต่อบอลลีแดงคงที่ จะยังคงทำให้ประสิทธิภาพในการประมาณค่าไม่เปลี่ยนแปลง และถ้าให้บอลลีทั้งหมดเป็นสีเดงนั่นคือ $\frac{w}{r}$ มีค่าเป็นศูนย์ ซึ่งทำให้เครื่องมือจะเหมือนวิธีการใช้คำถามโดยตรงและมีผลทำให้ความแปรปรวนในการประมาณค่ามีค่าต่ำสุดนั่นคือจาก $\hat{\text{Var}}(\hat{\pi}_i) = \left(\frac{r+w}{r} \right)^2 \left(\frac{1}{n} Y_i (1 - Y_i) \right)$ เมื่อ $w = 0$ แล้ว

$$\hat{\text{Var}}(\hat{\pi}_i) = \frac{1}{n} Y_i (1 - Y_i)$$
 นอกจากนี้จำนวนผู้ตอบ เป็นอีกตัวแปรหนึ่งที่มีผลต่อความแปรปรวน นั่นคือ ยิ่งเพิ่มจำนวนผู้ตอบมากยิ่งขึ้นทำให้การประมาณค่ามีประสิทธิภาพมากขึ้น ข้อดีของวิธีการตอบสนองการสุ่มนี้คือ สร้างความสบายใจแก่ผู้ตอบกล้าที่จะตอบความจริงมากยิ่งขึ้นทำให้ข้อมูลที่ได้น่าเชื่อถือมากยิ่งขึ้น และเครื่องมือที่ใช้จากโมเดลนี้สามารถใช้ได้ง่ายและในการคำนวณค่าประมาณพารามิเตอร์ง่ายกว่าวิธีเดิมเมื่อเปรียบเทียบกับวิธีหลิและคณะ [28] เมื่อนำเทคนิคดังกล่าวไปสำรวจ “ความบ่อยในการดูวิดีโอหนังลามกในรอบปี 2543 ของอาจารย์ นักศึกษาปริญญาโทภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยเชียงใหม่” พบว่าในรอบปี 2543 อาจารย์นักศึกษาปริญญาโทภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยเชียงใหม่ ไม่เคยดูหนังลามกเลยคิดเป็น 31.9% เคยดู 1 ครั้ง คิดเป็น 19.04% เคยดู 2 ครั้ง , 3 ครั้ง , 7 ครั้ง , 8 ครั้ง , 10 ครั้ง และมากกว่า 10 ครั้งคิดเป็น 6.19% เท่ากัน ซึ่งโดยรวม การเคยดูหนังลามกคิดเป็น 69.1% ของของอาจารย์ นักศึกษาปริญญาโทภาควิชาสถิติทั้งหมด

2.6 เอกสารที่เกี่ยวข้องอื่นๆ

เนื่องจากหลังจากศึกษาถึงวิธีการเทคนิคการตอบสนองการสุ่มทั้งหมดแล้ว ผู้วิจัยมีวัตถุประสงค์ที่จะนำวิธีการที่พัฒนาขึ้นแล้วเก็บข้อมูลจริงเพื่อศึกษาถึงสภาพปัญหาพฤติกรรมของเยาวชนไทยในสถาบันอาชีวศึกษาในอำเภอเมืองจังหวัดเชียงใหม่ เกี่ยวกับสภาพปัญหาการติดยาเสพติด และพฤติกรรมทางเพศ และการพนันฟุตบอล ซึ่งเป็นปัญหาที่รุนแรงอยู่ขณะนี้

เนื่องจากความเปลี่ยนแปลงทางเศรษฐกิจและสังคมอย่างรวดเร็ว ทำให้ปัญหาการแพร่ระบาดของยาเสพติดของเยาวชนในสถานศึกษามีแนวโน้มสูงขึ้น รวมทั้งชนิดของยาเสพติดก็มีความร้ายแรงขึ้นตามลำดับ จากสถิติการติดตามแก้ไขปัญหายาเสพติดในสถานศึกษาของฝ่ายสารเสพติด กองสารวัตรนักเรียน กรมพลศึกษา [3] พบว่านักเรียนระดับการศึกษามัธยมตอนต้นถึงระดับอาชีวศึกษา ส่วนมากใช้สารเสพติดเริ่มจากบุหรี่ กัญชา สารระเหย มอร์ฟีน แอมเฟตามีน จนถึงเฮโรอีน จากการสำรวจข้อมูลของสถาบันศึกษา นักเรียนที่ใช้สารเสพติดส่วนใหญ่คือ เหล้า บุหรี่ และรองลงมาคือเฮโรอีน ยาบ้าและกัญชา

สำนักงานคณะกรรมการป้องกันและปราบปรามยาเสพติด [17] ได้แบ่งยาเสพติดออกเป็น 4 ประเภท คือตามการออกฤทธิ์ ตามแหล่งที่มา ตามกฎหมาย และแบ่งตามองค์การอนามัยโลก ซึ่งแบ่งตามกฎหมายนั้นได้แยกออกเป็น 2 ประเภทคือ ถูกกฎหมายและผิดกฎหมาย ประเภทถูกกฎหมายมีการวางขายตามร้านทั่ว ๆ ไป ตัวอย่างได้แก่ ยาแก้ไอ น้ำคำ บุหรี่ เหล้า กาแฟ เป็นต้น สำหรับประเภทผิดกฎหมายจะเป็นยาเสพติดที่ให้โทษแก่ผู้เสพ ซึ่งแบ่งออกเป็นกลุ่มย่อยได้อีก คือกลุ่มยาเสพติดให้โทษเช่น มอร์ฟีน เฮโรอีน กระต่อม กัญชา กลุ่มวัตถุออกฤทธิ์ เช่น แอมเฟตามีน เซโคบาร์บิทัล และกลุ่มสารระเหย เช่น ทินเนอร์ กาว แล็กเกอร์

นงคราญ ณรงค์ศักดิ์ [8] ได้ศึกษาเรื่อง การดำเนินการป้องกันยาเสพติดของผู้บริหารโรงเรียนมัธยมศึกษา สังกัดกรมสามัญศึกษา จังหวัดเชียงราย ปี 2540 จำนวน 84 แห่งพบว่า สภาพการดำเนินการป้องกันยาเสพติดในโรงเรียนมีการสำรวจเกี่ยวกับการใช้ยาเสพติดของนักเรียน พบว่ามีการใช้บุหรี่ปากที่สุด รองลงมาคือ แอมเฟตามีน สารระเหย สุรา และสาเหตุของการแพร่เกินครึ่งนำมาเสพในโรงเรียน โดยเฉพาะในห้องน้ำ ห้องส้วมมีการมั่วสุมมากที่สุด

จากการวิเคราะห์แนวโน้มการใช้สารเสพติดจากอัตราการเข้ารับการบำบัดรักษารายใหม่ทุก 5 เดือน ในพ.ศ. 2533-2536 ของสำนักงานคณะกรรมการป้องกันและปราบปรามยาเสพติด [17] พบว่า ปี 2536 มีแนวโน้มของการใช้สารเสพติดในกลุ่มนักเรียน เพิ่มขึ้นจากปี 2533 เป็น 5 เท่า โดยปี 2533 มีนักเรียนเข้ารับการบำบัด 448 คน ปี 2534 มีจำนวน 697 คน ปี 2535 มีจำนวน 1,250 คน และในปี 2536 จำนวน 2,307 คน ซึ่งในจำนวนนี้ มีนักเรียนที่เรียนอยู่ในระดับมัธยมศึกษา มากที่สุด คือ ร้อยละ 45.52 รองลงมาคือระดับอาชีวศึกษา ร้อยละ 38.93

นัฏฐดี พลวัง [7] การศึกษาสถานการณ์ยาเสพติดในโรงเรียนมัธยมศึกษาเขตอำเภอเมืองเชียงใหม่ จังหวัดเชียงใหม่ ระหว่างปี 2541-2546 โดยใช้เทคนิคการสร้างภาพอนาคต พบว่า ยาเสพติดที่คาดว่าจะมีการแพร่ระบาดมากที่สุดในกลุ่มนักเรียนมัธยมศึกษา ได้แก่ บุหรี่ รองลงมา ยาบ้าและสารระเหย โดยบุหรี่และสารระเหย วิธีการแพร่โดยนักเรียนเป็นผู้นำเข้ามา ส่วนยาบ้า นั้นมีบุคคลภายนอกที่เป็นผู้ค้ายาบ้า เข้าไปทำความคุ้นเคยกับนักเรียนที่มีพฤติกรรมเบี่ยงเบน เช่น

หนีเรียน เทียวเตร โดยให้ทำการทดลองเสฟโดยไม่ต้องซื้อ แล้วเปลี่ยนการให้เป็นการซื้อ เมื่อต้องซื้อก็ไม่อยากเสียเงิน หรือกำลังซื้อไม่เพียงพอจะเปลี่ยนตัวเองมาเป็นผู้ค้าหรือหาลูกค้าจากโรงเรียน โดยลักษณะการเสฟ คาดว่าเป็นการสุุดคม และลักษณะการเสฟจะเสฟไปเรื่อย ๆ ไม่แน่นอนตามลักษณะของยาเสฟติดและความนิยม

มานพ วรรณวงศ์ [11] ได้ทำการศึกษาเรื่องกิจกรรมการป้องกันการใช้สารเสพติดของนักเรียนโรงเรียนศรีธนาพาณิชยการเทคโนโลยีเชียงใหม่ โดยศึกษานักเรียน 234 คน ซึ่งเป็นนักเรียนระดับประกาศนียบัตรวิชาชีพ (ปวช.) ชั้นปีที่ 1-3 โดยใช้แบบสอบถามโดยตรงเป็นเครื่องมือพบว่าสภาพปัญหาคือนักเรียน ร้อยละ 62.82 ไม่เคยใช้สารเสฟติดเลยร้อยละ 37.18 เคยใช้สารเสฟติดโดยเคยเสฟสุรามากที่สุดถึงร้อยละ 31.19 และรองลงมาคือ บุหรี่ ยาบ้า กัญชา และสารระเหย คิดเป็นร้อยละ 10.68, 6.41, 2.56 และ 0.85 ตามลำดับ

นอกจากปัญหายาเสฟติดแล้ว ปัญหาเกี่ยวกับพฤติกรรมทางเพศยังเป็นปัญหาที่รุนแรงในขณะนี้ เช่น ปัญหาการมีเพศสัมพันธ์ในวัยเรียน การมีเพศสัมพันธ์ในเพศเดียวกัน การขายบริการทางเพศของนักเรียนนักศึกษา ทำให้เกิดปัญหาการทำแท้งขึ้นด้วยความรู้เท่าไม่ถึงการณ์ ปัญหาเหล่านี้งานวิจัยในเชิงปริมาณค่อนข้างน้อยทั้งนี้เป็นเพราะ เป็นปัญหาที่ค่อนข้างละเอียดอ่อนโดยเฉพาะค่านิยมของสังคมไทย การไม่ยอมรับเรื่องเพศ ทั้ง ๆ ที่ปัญหาเหล่านี้ทวีคูณขึ้นเรื่อย ๆ ในหมู่เยาวชนอันเนื่องมาจากความก้าวหน้าของเทคโนโลยี กระแสของโลกาภิวัตน์ ผลักดันให้ประเทศไทยเกิดการเปลี่ยนแปลงในด้านต่าง ๆ ทั้งด้านเศรษฐกิจ สังคมวัฒนธรรม วิทยาศาสตร์และเทคโนโลยี ตลอดจนการเมืองการปกครองการนำข่าวสารทั้งที่เป็นความรู้และวัฒนธรรมจากทั่วโลก เข้าสู่การรับรู้ของสังคมไทยอย่างรวดเร็วและต่อเนื่อง เป็นผลทำให้คนในสังคมหนึ่งเกิดความไม่เอียงที่จะประพฤติปฏิบัติตามวัฒนธรรมของอีกสังคมหนึ่ง ส่งเสริมให้เกิดการเลียนแบบซึ่งกันและกันอย่างรวดเร็ว รวมถึงความสัมพันธ์ระหว่างเพศ ซึ่งปัจจุบันได้เปลี่ยนแปลงไปมาก ในสมัยก่อนสังคมไทยกำหนดให้หญิงไทยรักษาวลสงวนตัว หญิงชายไม่ถูกเนื้อต้องตัวกันก่อนแต่งงาน และไม่อนุญาตให้เด็กในวัยเรียนได้คบเพื่อนต่างเพศจนกว่าจะถึงวัยอันสมควร การรู้จักเพศตรงข้ามต้องอยู่ภายใต้การดูแลของบิดามารดาหรือผู้ปกครอง การมีเพศสัมพันธ์ก่อนแต่งงานนั้นเป็นเรื่องน่าอายแต่ในปัจจุบันชายหญิงมีอิสระมากขึ้น ในการคบค้าสมาคมกับเพื่อนต่างเพศ การแสดงความสนิทสนมจะมีมากกว่าคนรุ่นก่อน การแสดงความสัมพันธ์ระหว่างชายหญิงที่เป็นคู่รักกันมีแนวโน้มที่จะเปิดเผยมากยิ่งขึ้น เช่น เดินจู๋มือกันโอบกอดไปตามถนน ป้อนอาหารกันในร้านอาหาร นอนหนุนตักกันในสวนสาธารณะ เป็นต้น กลุ่มคนวัยหนุ่มสาวมีแนวโน้มที่จะยอมรับและถือปฏิบัติรูปแบบความสัมพันธ์ระหว่างเพศดังกล่าวอย่างรวดเร็วและมากกว่ากลุ่มอื่นเพราะวัยนี้มีความอยากรู้อยากลอง ส่วนหนึ่งเนื่องจากแรงขับทางเพศของเขาเหล่านั้น

สุพัตรา สุภาพ [16] กล่าวถึงวัยรุ่นหนุ่มสาวนับเป็นช่วงที่สำคัญต่อชีวิตมาก เพราะยังอยู่ในระหว่างการศึกษาเล่าเรียนเพื่อสร้างพื้นฐานให้แก่อนาคต หากเขาเหล่านั้นเกิดความผิดพลาดในเรื่องการมีเพศสัมพันธ์ จนเป็นเหตุทำให้เกิดความผิดหวังต่อแท้ ก็จะทำให้เกิดการหักเหของชีวิตนำไปสู่ความสิ้นหวังได้โดยง่าย กลุ่มหนุ่มสาวที่อยู่ในระดับอุดมศึกษาเป็นกลุ่มที่มีโอกาสในด้านการมีความสัมพันธ์ระหว่างเพศได้ง่าย เพราะมีอิสระในการพบปะสมาคมกับเพื่อนต่างเพศมากกว่ากลุ่มอื่น ๆ และจะมีความเป็นไปได้สูงมากขึ้นที่จะนำไปสู่การมีความสัมพันธ์กับเพศตรงข้ามในที่สุดอาจนำไปสู่การมีเพศสัมพันธ์ก่อนแต่งงาน ถ้าไม่ควบคุมความต้องการทางเพศได้อย่างเหมาะสมอาจจะก่อให้เกิดอุปสรรคต่อการเรียน และเกิดปัญหาอื่น ๆ ตามมา การเที่ยวเตร่มั่วสุมในสถานเริงรมย์ต่าง ๆ นำไปสู่การมีเพศสัมพันธ์ในกลุ่มเพื่อนด้วยตนเอง เกิดการตั้งครรภ์ หรือติดโรคทางเพศสัมพันธ์ เช่น กามโรค และเอดส์ เป็นต้น ซึ่งล้วนเป็นปัญหาสุขภาพและภาระทางสังคมทั้งสิ้น นักเรียนในวัยนี้เป็นวัยที่มีการพัฒนาทางด้านเพศถึงระดับหนึ่งแต่ยังขาดความรับผิดชอบจึงเป็นกลุ่มเสี่ยงต่อการมีเพศสัมพันธ์กับเพื่อนต่างเพศ และพบว่าในกลุ่มที่พักอยู่กับพ่อแม่มีพฤติกรรม ดังกล่าวเพียงร้อยละ 10 ส่วนกลุ่มที่ไม่ได้พักอยู่กับพ่อแม่และมีพฤติกรรมข้างต้นถึงร้อยละ 60 ทั้งนี้เพราะอิทธิพลของเพื่อน โอกาสที่จะทำให้นักศึกษาในวัยนี้ขาดความยับยั้งชั่งใจเป็นไปได้ง่าย หากไม่ได้รับการดูแลอย่างใกล้ชิดและความอบอุ่นที่เพียงพอจากบิดามารดาหรือผู้ใหญ่ นอกจากนี้ผลการสำรวจของสถาบันเอแบคโกลด์อินเตอร์เน็ต เมื่อปี พ.ศ. 2540 จากกลุ่มตัวอย่างในการศึกษาเป็นนักเรียนนักศึกษาชาย ที่เคยมีเพศสัมพันธ์มาแล้วในเขตกรุงเทพมหานคร ขอนแก่น เชียงใหม่ และสงขลา ซึ่งกำลังศึกษาในสายอาชีพระดับประกาศนียบัตรวิชาชีพ (ปวช.) ประกาศนียบัตรวิชาชีพชั้นสูง (ปวส.) ระดับปริญญาตรี และกำลังศึกษาสายสามัญระดับมัธยมศึกษาตอนปลาย พบว่าร้อยละ 72.1 ของตัวอย่าง มีเจตคติเห็นด้วยต่อการมีเพศสัมพันธ์กับคู่รักโดยยังไม่แต่งงานร้อยละ 44.9 เห็นด้วยต่อการมีเพศสัมพันธ์กับเพื่อนร้อยละ 46.2 ระบุว่าเคยมีเพศสัมพันธ์กับเพื่อนร้อยละ 44.8 ระบุว่าเคยมีเพศสัมพันธ์กับเพื่อนต่างสถาบัน ร้อยละ 45.8 ระบุว่าไม่ได้ใช้ถุงยางอนามัยป้องกันเมื่อมีเพศสัมพันธ์กับคู่รัก กลุ่มนักเรียนโรงเรียนอาชีวศึกษาเหล่านี้มีอายุประมาณ 15-19 ปี ซึ่งอยู่ในระยะวัยรุ่น เป็นวัยที่มีการเปลี่ยนแปลงอย่างรวดเร็ว ทั้งทางด้านร่างกายและจิตใจ นักเรียนกลุ่มนี้จะมีพัฒนาการทางด้านเพศถึงระดับหนึ่ง แต่ยังขาดความรับผิดชอบ จึงจัดได้ว่านักเรียนในระดับนี้ เป็นกลุ่มเสี่ยงต่อการมีเพศสัมพันธ์กับเพื่อนต่างเพศ

จริยา พงศ์วิวัฒน์ [4] ได้ศึกษาทัศนคติในเรื่อง การเลือกคู่ครอง และการแต่งงานของนักศึกษาปีที่ 1 มหาวิทยาลัยธรรมศาสตร์ พบว่า นักศึกษาส่วนใหญ่ยอมรับพฤติกรรมการจับมือถือแขนของคู่รักว่าไม่ใช่สิ่งเสียหาย โดยนักศึกษาชายมีคะแนนเฉลี่ยมากกว่านักศึกษาหญิง และนักศึกษาทั้งหมดเห็นด้วยกับทัศนคติที่ว่า เป็นเรื่องธรรมดาที่ผู้ชายจะมีประสบการณ์ทางเพศก่อน

แต่งงาน และเป็นเรื่องเลื่อมเสียที่ผู้หญิงจะมีประสบการณ์ทางเพศก่อนแต่งงาน และในทัศนคติดังกล่าวนักศึกษาหญิงเห็นด้วยมากกว่านักศึกษาชาย

จิราภรณ์ อารยะรังสฤษฎ์ [5] ได้ทำการศึกษาทัศนคติเชิงจริยธรรมของนักศึกษามหาวิทยาลัยเชียงใหม่ เกี่ยวกับความสัมพันธ์ระหว่างชายหญิงและรักร่วมเพศ ของนักศึกษามหาวิทยาลัยเชียงใหม่ พบว่านักศึกษามีทัศนคติเชิงจริยธรรมที่มีต่อความสัมพันธ์ระหว่างชาย-หญิง และความสัมพันธ์ระหว่างพวกเรารักร่วมเพศในสถานการณ์ต่าง ๆ มีความแตกต่าง โดยนักศึกษาจะให้การยอมรับพฤติกรรมการนัดพบกันอย่างเปิดเผยมากที่สุด และยอมรับการกอดจูบก้นในสวนสาธารณะน้อยที่สุด ไม่ว่าจะเป็นความสัมพันธ์แบบรักร่วมเพศก็ตาม

วรวิทย์ ตาปิง [13] ได้ศึกษาความคิดเห็นของอาจารย์และนักศึกษาในสถาบันอุดมศึกษา จังหวัดเชียงใหม่ และพฤติกรรมของนักศึกษาเกี่ยวกับความสัมพันธ์ระหว่างเพศ พบว่านักศึกษาชายและนักศึกษาหญิงความคิดเห็นด้านการสมาคมกับเพื่อนต่างเพศแตกต่างกัน โดยมีค่าเฉลี่ยของนักศึกษาชายสูงกว่าค่าเฉลี่ยของนักศึกษาหญิงในประเด็นที่ว่าในวัยเรียนควรมีนัดกับเพื่อนต่างเพศ การถูกเนื้อต้องตัวของคู่รักเป็นเรื่องธรรมดา การสมาคมกับเพื่อนต่างเพศมากกว่าหนึ่งคนในเวลาเดียวกันเป็นสิ่งไม่เสียหาย หญิงมีคู่รักมากกว่าหนึ่งคนเป็นสิ่งเสียหาย และชายหญิงควรมีประสบการณ์ทางเพศก่อนแต่งงาน

ศรีเกษ ัญญาวินิชกุล [14] ได้ทำการศึกษาเรื่องเหตุผลในการตัดสินใจเกี่ยวกับพฤติกรรมทางเพศของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ในเขตเมืองเชียงใหม่ พบว่านักเรียนชายกับนักเรียนหญิงที่ตัดสินใจไม่มีเพศสัมพันธ์ ใช้เหตุผลด้านค่านิยมประกอบการตัดสินใจเกี่ยวกับพฤติกรรมทางเพศแตกต่างกัน โดยจะเห็นว่านักเรียนหญิงที่ตัดสินใจไม่มีเพศสัมพันธ์ให้ความสำคัญ กับพรหมจรรย์ของเพศหญิงว่า เป็นสิ่งสำคัญสำหรับการแต่งงานมากกว่านักเรียนชาย นักเรียนชายกับนักเรียนหญิงที่ตัดสินใจไม่มีเพศสัมพันธ์ใช้เหตุผลด้านจริยธรรมประกอบการตัดสินใจ เกี่ยวกับพฤติกรรมทางเพศแตกต่างกัน โดยจะเห็นว่า นักเรียนหญิงตัดสินใจไม่มีเพศสัมพันธ์ ให้เหตุผลว่าหญิงและชายไม่ควรจะมีเพศสัมพันธ์ในวัยเรียน ถือว่าเป็นการผิดศีลธรรมและยังทำให้ฝ่ายหญิงเลื่อมเสียชื่อเสียง นักเรียนจะคำนึงถึงหน้าที่ของตนเองในสังคมคือการศึกษาเล่าเรียน ตลอดจนมองถึงผลเสียที่ตามมาจากการมีเพศสัมพันธ์ในวัยเรียนมากกว่านักเรียนชาย นักเรียนชายกับนักเรียนหญิงที่ตัดสินใจไม่มีเพศสัมพันธ์ ใช้เหตุผลด้านสาธารณสุขประกอบการตัดสินใจเกี่ยวกับพฤติกรรมทางเพศแตกต่างกัน โดยจะเห็นว่านักเรียนหญิงที่ตัดสินใจไม่มีเพศสัมพันธ์จะคำนึงถึงผลเสียที่ตามมาจากการมีเพศสัมพันธ์มากกว่านักเรียนชาย