บทที่ 2

ทฤษฎีที่เกี่ยวข้องกับการวิจัย

ในการวิจัยครั้งนี้สิ่งที่สนใจศึกษาคือ การเปรียบเทียบวิธีการประมาณค่าผลต่างค่าสัดส่วนของสองประชากรที่มีการแจกแจงแบบแบร์นูลลีและเป็นอิสระกัน ซึ่งประกอบด้วย วิธีการประมาณแบบฉบับ วิธีการประมาณของ Newcombe ซึ่งมีแนวคิดมาจากวิธีการประมาณค่าสัดส่วนด้วยรากของสมการกำลังสอง วิธีการประมาณของ Jeffrey ซึ่งนำแนวความคิดของเบส์มาประยุกต์ใช้ และวิธีการประมาณแบบจัดค่ากลางใหม่ ทั้งนี้การศึกษาเปรียบเทียบ จะพิจารณาจากค่าสัมประสิทธิ์ความเชื่อมั่นของช่วงความเชื่อมั่น และค่าความยาวเฉลี่ยของช่วงความเชื่อมั่นที่คำนวณได้จากแต่ละวิธี ดังรายละเอียดในวัตถุประสงค์ของการวิจัยที่ได้กล่าวไปแล้วในบทที่ 1 ในบทนี้จะกล่าวถึงรายละเอียดของแต่ละวิธีการประมาณ ตลอดจนทฤษฎีอื่นๆที่เกี่ยวข้องกับการวิจัย โดยมีรายละเอียดต่างๆดังนี้

2.1 ทฤษฎีบทลิมิตเข้าสู่ส่วนกลาง

เมื่อสุ่มตัวอย่างขนาด n หน่วย จากประชากรที่มีการแจกแจงแบบใดๆ โดยที่ประชากรดังกล่าวมีค่าเฉลี่ยเท่ากับ $E(X)$ และค่าความแปรปรวนเท่ากับ $Var(X)$ ซึ่งค่าทั้งสองเป็นค่าจำกัดในกรณีที่ขนาดตัวอย่างมีขนาดใหญ่มากพอ การแจกแจงของค่าเฉลี่ยตัวอย่างจะมีการแจกแจงลู่เข้าสู่การแจกแจงแบบปกติ ที่มีค่าเฉลี่ยเท่ากับ $E(X)$ และค่าความแปรปรวนเท่ากับ $Var(X)/n$

กำหนดให้ $X_1, X_2, \ldots\ldots, X_n$ เป็นตัวแปรสุ่มที่มีการแจกแจงเหมือนกันและเป็นอิสระกัน โดยการแจกแจงดังกล่าวมีค่าเฉลี่ยเท่ากับ $E(X)$ และความแปรปรวนเท่ากับ $Var(X)$ ซึ่งค่าทั้งสองมีค่าจำกัด

ให้ $\overline{X}$ เป็นค่าเฉลี่ยของตัวอย่างสุ่ม มีค่า $= (X_1 + X_2 + \ldots\ldots + X_n)/n$
ในกรณีที่ $n \to \infty$ จะได้ว่า

$$\overline{X} \sim N(E(X), Var(X)/n) \qquad \text{(โดยประมาณ)}$$

หรือ

$$Z = \frac{\overline{X} - E(X)}{\sqrt{Var(X)/n}} \sim N(0,1) \qquad \text{(โดยประมาณ)}$$

2.2 การแจกแจงแบร์นูลลี และการแจกแจงทวินาม [1]

ในหลายการทดลองสุ่ม จะมีผลลัพธ์ที่เราจำแนกออกได้เป็นสองผลลัพธ์ หรือสองเหตุการณ์ เช่น เห็นด้วยหรือไม่เห็นด้วย ใช้การได้หรือใช้การไม่ได้ ผ่านหรือไม่ผ่าน เลขคู่หรือเลขคี่ ชายหรือหญิง และรักษาหายหรือรักษาไม่หาย เป็นต้น ถ้าให้ X แทนผลลัพธ์ของการทดลองสุ่ม ดังนั้น X จะเป็นตัวแปรสุ่มมีค่าเป็นไปได้สองค่า

ตัวแปรสุ่ม X มีการแจกแจงสองจุด (Two – point Distribution) ถ้า X มีฟังก์ชันความน่าจะเป็น

$$p(x) = \begin{cases} p & \text{ถ้า } x = x_1 \\ q & \text{ถ้า } x = x_2 \\ 0 & \text{อื่นๆ} \end{cases}$$

ซึ่ง $q = 1 - p$, $0 \le p \le 1$

ถ้าให้ $x_1 = 1$ และ $x_2 = 0$ หรือกลับกันเราเรียก X ว่า ตัวแปรสุ่มแบร์นูลลี (Bernoulli random variable) หรือ X มีการแจกแจงแบร์นูลลี เขียนแทนด้วย $X \sim b(1, p)$ และ X มีฟังก์ชันความน่าจะเป็นสำหรับ $x_1 = 1$ และ $x_2 = 0$ ดังนี้

$$p(x) = p(x;p) = \begin{cases} p^x q^{1-x} & x = 0,1 \\ 0 & x \ne 0,1 \end{cases}$$

โดยทั่วไป เราให้ $X = 1$ ถ้าผลลัพธ์ของการทดลองเป็นผลลัพธ์ในความต้องการ เรียกว่า เกิดผลสำเร็จ (Success) และให้ $X = 0$ ถ้าผลลัพธ์ไม่ตรงตามความต้องการ เรียกว่า ไม่เกิดผลสำเร็จ (Failure) เช่น $X = 1$ ถ้ามีความคิดเห็นด้วย และให้ $X = 0$ ถ้ามีความคิดไม่เห็นด้วย

$$E(X) = \sum_{x=0}^{1} x p^x q^{1-x} = (0)q + (1)p = p$$

$$Var(X) = E(X^2) - [E(X)]^2 = (0)^2 q + (1)^2 p - p^2 = pq$$

$$M_X(t) = E(e^{tX}) = e^0 q + e^t p = q + pe^t , -\infty < t < \infty$$

[1] มานพ วราภักดิ์, ทฤษฎีความน่าจะเป็น (กรุงเทพฯ: สำนักพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย, 2548), หน้า 43.

$$\psi_X(t) = E(e^{tX}) = e^0 q + e^t p = q + pe^t, \; -\infty < t < \infty$$

การทดลองที่มีสองผลลัพธ์เมื่อทำการทดลองซ้ำๆ n ครั้งอย่างเป็นอิสระกัน ซึ่งเรียกว่า การทดลองแบบแบร์นูลลี (Bernoulli Trial) ซ้ำกัน n ครั้งอย่างเป็นอิสระกัน แต่ละครั้งมีความน่าจะเป็นที่ผลลัพธ์เป็น 1 (เกิดผลสำเร็จ) เท่ากับ p และมีความน่าจะเป็นที่ผลลัพธ์เป็น 0 (ไม่เกิดผลสำเร็จ) เท่ากับ $q = 1 - p$ ถ้าให้ Y แทนจำนวนครั้งได้ผลสำเร็จในการทดลอง n ครั้ง

Y เป็นตัวแปรสุ่มทวินาม (Binomial Random Variable) แสดงว่า Y มีการแจกแจงแบบทวินาม (Binomial Distribution) โดยมีพารามิเตอร์คือ n และ p และเขียนแทนด้วย $Y \sim b(n, p)$ ซึ่งจะได้ฟังก์ชันความน่าจะเป็นของ Y คือ

$$P(Y = y) = \binom{n}{y} p^y q^{n-y} \qquad \text{สำหรับ } y = 0,1,\ldots\ldots,n \text{ และ } 0 \le p \le 1$$

โดยมีค่าคาดหวังและค่าความแปรปรวนเท่ากับ

$$\begin{aligned}
E(Y) &= \sum_{y=0}^{n} y \binom{n}{y} p^y (1-p)^{n-y} \\
&= \sum_{y=1}^{n} \frac{n(n-1)!}{(y-1)!(n-y)!} p^y (1-p)^{n-y} \\
&= (np) \sum_{y-1=0}^{n-1} \binom{n-1}{y-1} p^{y-1} (1-p)^{(n-1)-(y-1)} \\
&= np(p+q)^{n-1}, \; q = 1-p \\
&= np \quad (\text{เนื่องจากว่า } p + q = 1)
\end{aligned}$$

$$\begin{aligned}
Var(Y) &= E(Y^2) - [E(Y)]^2 \\
&= E(Y(Y-1)) + E(Y) - [E(Y)]^2
\end{aligned}$$

จาก

$$\begin{aligned}
E(Y(Y-1)) &= \sum_{y=0}^{n} y(y-1) \binom{n}{y} p^y (1-p)^{n-y} \\
&= \sum_{y=2}^{n} \frac{n(n-1)(n-2)!}{(y-2)!(n-y)!} p^y (1-p)^{n-y} \\
&= n(n-1)p^2 \sum_{y-2=0}^{n-2} \frac{(n-2)!}{(y-2)!(n-y)!} p^{y-2} (1-p)^{(n-2)-(y-2)}
\end{aligned}$$

$$= n(n-1)p^2 \sum_{y-2=0}^{n-2} \binom{n-2}{y-2} p^{y-2}(1-p)^{(n-2)-(y-2)}$$

$$= n(n-1)p^2 (p+q)^{n-2} \text{ , } q = 1-p$$

$$= n(n-1)p^2$$

ดังนั้น $\quad Var(Y) = n(n-1)p^2 + np - (np)^2$

$$= n^2p^2 - np^2 + np - n^2p^2$$

$$= np(1-p) = npq$$

2.3 ทฤษฎีการประมาณค่าพารามิเตอร์แบบช่วง

ในการประมาณค่าพารามิเตอร์โดยใช้วิธีการประมาณแบบจุด ซึ่งเป็นวิธีการที่ใช้ค่าประมาณเพียงค่าเดียวที่คำนวณได้จากตัวอย่างสุ่ม เป็นค่าประมาณสำหรับพารามิเตอร์ของประชากร เราไม่สามารถบอกได้ว่าค่าประมาณแบบจุดนั้น มีค่าเท่ากับค่าพารามิเตอร์หรือแตกต่างจากค่าพารามิเตอร์อย่างไร จึงได้มีแนวคิดที่จะนำเอาความแปรปรวนของตัวประมาณแบบจุด การแจกแจงตัวอย่างของตัวประมาณ และค่าประมาณแบบจุดมาประกอบเข้าด้วยกันเพื่อสร้างช่วงความเชื่อมั่นขึ้นมา และอาศัยการแจกแจงของตัวประมาณเป็นตัวกำหนดความมั่นใจได้ว่าช่วงความเชื่อมั่นที่สร้างขึ้นจะคลอบคลุมค่าพารามิเตอร์ที่แท้จริงด้วยความน่าจะเป็นเท่าใด

ให้ $X_1, X_2, \ldots\ldots, X_n$ เป็นตัวอย่างสุ่มจากการแจงแจงซึ่งมี θ เป็นพารามิเตอร์ที่ไม่ทราบค่า

สมมติสามารถหาตัวสถิติ $L(X_1, X_2, \ldots\ldots, X_n)$ และ $U(X_1, X_2, \ldots\ldots, X_n)$ ซึ่งสำหรับค่าจริง θ ใดๆ

$$P(L(X_1, X_2, \ldots\ldots, X_n) < \theta < U(X_1, X_2, \ldots\ldots, X_n)) \qquad = 1-\alpha$$

โดยที่ความน่าจะเป็น $1-\alpha$ เป็นค่าคงที่ $(0 < \alpha < 1)$

จากนี้เมื่อทราบค่าของ $X_i = x_i (i = 1, \ldots\ldots, n)$ และค่าของ $L(X_1, X_2, \ldots\ldots, X_n)$ และ $U(X_1, X_2, \ldots\ldots, X_n)$ สมมติให้เป็น l และ u ตามลำดับ ดังนั้นจะได้ช่วง (l, u) และเรียกช่วง (l, u) ว่า ช่วงความเชื่อมั่น $100(1-\alpha)$ เปอร์เซ็นต์สำหรับ θ (100(1- α)% Confidence Interval for θ) และเรียกค่า l ว่า ขีดจำกัดความเชื่อมั่นล่าง (Low Confidence Limit) เรียกค่า

u ว่า ขีดจำกัดความเชื่อมั่นบน (Upper Confidence Limit) และเรียกค่า $(1-\alpha)$ ว่า สัมประสิทธิ์ความเชื่อมั่น (Confidence Coefficient)

2.4 การแจกแจงของผลต่างของค่าสัดส่วนตัวอย่าง

ในการทดลองสุ่มใดๆก็ตามที่มีผลลัพธ์ที่เป็นไปได้ 2 อย่าง คือ เกิดสิ่งที่สนใจ หรือผลสำเร็จ (Success) และ เกิดสิ่งที่ไม่สนใจ หรือผลล้มเหลว (Failure) ด้วยค่าความน่าจะเป็น p และ $q = 1 - p$ ตามลำดับ การแจกแจงซึ่งอธิบายการทดลองลักษณะนี้เรียกว่า การแจกแจงความน่าจะเป็นแบบแบร์นูลลี (Bernoulli Probability Distribution) ด้วยพารามิเตอร์ p

$$X \sim Ber(p)$$

และฟังก์ชันความน่าจะเป็นจะอยู่ในรูป

$$f(x;p) = p^x(1-p)^{1-x},\ x = 0,1 \text{ และ } 0 \le p \le 1$$

ค่าความน่าจะเป็น p คือค่าสัดส่วนประชากรที่เกิดผลสำเร็จ

สำหรับพารามิเตอร์ p หาตัวประมาณของ p ด้วยวิธีความควรจะเป็นสูงสุด (Maximum Likelihood Method) ดังนี้

กำหนดให้ $X_1X_2,.....,X_n$ เป็นตัวแปรสุ่มอิสระกันมีการแจกแจงแบบแบร์นูลลี โดยมีฟังก์ชันความน่าจะเป็นจะอยู่ในรูป

$$f(x;p) = p^x(1-p)^{1-x},\ x = 0,1 \text{ และ } 0 \le p \le 1$$

ได้ฟังก์ชันความควรจะเป็นคือ

$$L(p;x) = \prod_{i=1}^{n} p^{x_i}(1-p)^{x_i}$$

$$= p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i}$$

$$\ln L(p;x) = \sum_{i=1}^{n} x_i \ln p + \left(n - \sum_{i=1}^{n} x_i\right)\ln(1-p)$$

$$\frac{\partial \ln L(p;x)}{\partial p} = \frac{\sum_{i=1}^{n} x_i}{p} - \frac{\left(n - \sum_{i=1}^{n} x_i\right)}{1-p} = 0$$

$$\frac{\sum_{i=1}^{n} x_i}{p} = \frac{n - \sum_{i=1}^{n} x_i}{1-p}$$

$$\sum_{i=1}^{n} x_i - p\sum_{i=1}^{n} x_i = np - p\sum_{i=1}^{n} x_i$$

$$p = \frac{\sum_{i=1}^{n} x_i}{n}$$

ดังนั้น $\hat{p} = \frac{\sum_{i=1}^{n} X_i}{n}$ เป็นตัวประมาณความควรจะเป็นสูงสุด (Maximum Likelihood Estimator) ของ p

กำหนดให้ $\{X_{11}, X_{12}, \ldots, X_{1n_1}\}$ เป็นตัวอย่างสุ่มขนาด n_1 จากการแจกแจงแบบแบร์นูลลี $Ber(p_1)$ และ $\{X_{21}, X_{22}, \ldots, X_{2n_2}\}$ เป็นตัวอย่างสุ่มขนาด n_2 จากการแจกแจงแบบแบร์นูลลี $Ber(p_2)$ โดยที่ค่า p_1 และ p_2 เป็นค่าความน่าจะเป็นที่เกิดผลสำเร็จในประชากรที่1 และ 2 ตามลำดับ หรือค่าสัดส่วนประชากรของประชากรที่ 1 และ 2 ตามลำดับ

จะได้ $Y_1 = \sum_{j=1}^{n_1} X_{1j}$ มีการแจกแจงทวินามเขียนแทนด้วย $Y_1 \sim B(n_1, p_1)$ มีค่าเฉลี่ยเท่ากับ $n_1 p_1$ และค่าความแปรปรวนเท่ากับ $n_1 p_1 q_1$, $q_1 = 1 - p_1$

และ $Y_2 = \sum_{j=1}^{n_2} X_{2j}$ มีการแจกแจงทวินามเขียนแทนด้วย $Y_2 \sim B(n_2, p_2)$ มีค่าเฉลี่ยเท่ากับ $n_2 p_2$ และค่าความแปรปรวนเท่ากับ $n_2 p_2 q_2$, $q_2 = 1 - p_2$

ในการประมาณค่าสัดส่วนประชากร (p_1, p_2) โดยใช้ค่าสัดส่วนตัวอย่าง $(\hat{p}_1, \hat{p}_2)$ โดยที่ $\hat{p}_1 = \frac{Y_1}{n_1}$ และ $\hat{p}_2 = \frac{Y_2}{n_2}$ เมื่อพิจารณาค่าเฉลี่ยของ $\hat{p}_1$ และ $\hat{p}_2$

$$E(\hat{p}_i) = E\left(\frac{Y_i}{n_i}\right) = \frac{1}{n_i} E(Y_i) = \frac{n_i p_i}{n_i} = p_i \quad , i = 1,2$$

แสดงให้เห็นว่า $\hat{p}_1$ และ $\hat{p}_2$ เป็นตัวประมาณที่ไม่เอนเอียง (Unbiased Estimator) ของ p_1 และ p_2 ส่วนค่าความแปรปรวนของ $\hat{p}_1$ และ $\hat{p}_2$ คือ

$$Var(\hat{p}_i) = Var\left(\frac{Y_i}{n_i}\right) = \frac{1}{n_i^2} Var(Y_i) = \frac{1}{n_i^2}(n_i p_i q_i) = \frac{p_i q_i}{n_i} \quad , q_i = 1 - p_i$$

สำหรับตัวแปรสุ่ม $\hat{p}_1$ และ $\hat{p}_2$ นี้ไม่ได้มีการแจกแจงในรูปแบบที่ใช้ประโยชน์ง่าย อย่างไรก็ตาม ถ้า n_1 และ n_2 มีขนาดใหญ่ จากทฤษฎีบทขีดจำกัดส่วนกลาง (Central Limit Theorem) จะได้ว่า $\hat{p}_1$ และ $\hat{p}_2$ มีการแจกแจงแบบปกติโดยประมาณ มีค่าเฉลี่ยเท่ากับ p_i และความแปรปรวนเท่ากับ $\frac{p_i q_i}{n_i}$ เมื่อ $q_i = 1 - p_i; i = 1,2$

นั่นคือ เมื่อ $n_i \to \infty$

$$\hat{p}_i \sim N\left(p_i, \frac{p_i q_i}{n_i}\right) \quad \text{โดยประมาณ}$$

$$Z = \frac{\hat{p}_i - p_i}{\sqrt{\frac{p_i(1-p_i)}{n_i}}} \sim N(0,1) \quad \text{โดยประมาณ}$$

เนื่องจาก $\{X_{11}, X_{12}, \ldots, X_{1n_1}\}$ และ $\{X_{21}, X_{22}, \ldots, X_{2n_2}\}$ เป็นตัวอย่างสุ่ม 2 ชุดจากการแจกแจงแบบ $Ber(p_1)$ และ $Ber(p_2)$ ตามลำดับที่เป็นอิสระกัน เมื่อ $Y_1 = \sum_{j=1}^{n_1} X_{1j}$ และ $Y_2 = \sum_{j=1}^{n_2} X_{2j}$ เป็นตัวแปรสุ่มที่มีการแจกแจงแบบ $B(n_1, p_1)$ และ $B(n_2, p_2)$ ตามลำดับที่เป็นอิสระกัน เมื่อ $\hat{p}_1 = \frac{Y_1}{n_1}$ และ $\hat{p}_2 = \frac{Y_2}{n_2}$ จะได้ว่า

$$E(\hat{p}_1 - \hat{p}_2) = p_1 - p_2 \quad \text{และ} \quad Var(\hat{p}_1 - \hat{p}_2) = \frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}$$

โดยทฤษฎีบทลิมิตสู่ส่วนกลาง เมื่อขนาดตัวอย่าง n_1 และ n_2 มีขนาดใหญ่ จะได้ว่า $\hat{p}_1 - \hat{p}_2$ มีการแจกแจงแบบปกติโดยประมาณ มีค่าเฉลี่ยเท่ากับ $p_1 - p_2$ และค่าความแปรปรวนเท่ากับ $\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}$ นั่นคือ เมื่อ n_1 และ $n_2 \to \infty$

$$\hat{p}_1 - \hat{p}_2 \sim N\left(p_1 - p_2, \frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}\right) \quad \text{โดยประมาณ}$$

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}} \sim N(0,1) \quad \text{โดยประมาณ}$$

2.5 วิธีการประมาณค่าแบบช่วงสำหรับผลต่างค่าสัดส่วนสองประชากร

ในการประมาณค่าแบบช่วงของค่าผลต่างระหว่างสัดส่วนของสองประชากร $(p_1 - p_2)$ เมื่อกำหนดความน่าจะเป็นของความผิดพลาด α หรือสัมประสิทธิ์ความเชื่อมั่น $1-\alpha$ (Confidence Coefficient) และเมื่อตัวอย่างสุ่มมีขนาดใหญ่ โดยทฤษฎีบทขีดจำกัดส่วนกลาง $p_1 - p_2$ จะมีการแจกแจงแบบปกติโดยประมาณ มีค่าเฉลี่ยเป็น $p_1 - p_2$ และค่าความแปรปรวนเป็น $\left[\frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}\right]$ จะได้

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}}} \sim N(0,1) \text{ โดยประมาณ}$$

ดังนั้น
$$\Pr(-Z_{1-\alpha/2} < Z < Z_{1-\alpha/2}) \approx 1-\alpha$$

$$\Pr(-Z_{1-\alpha/2} < \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}}} < Z_{1-\alpha/2}) \approx 1-\alpha$$

จะได้ช่วงความเชื่อมั่น $(1-\alpha)100\%$ โดยประมาณสำหรับ $p_1 - p_2$ คือ

$$(\hat{p}_1 - \hat{p}_2) \pm Z_{1-\alpha/2}\sqrt{\frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}}$$

การประมาณค่าแบบช่วงสำหรับค่าผลต่างระหว่างค่าสัดส่วนของประชากร 2 กลุ่มในงานวิจัยครั้งนี้ จะใช้วิธีประมาณดังนี้

1. วิธีการประมาณแบบฉบับ (The Classical Method)

จากรูปแบบของการประมาณค่าแบบช่วงสำหรับผลต่างระหว่างค่าสัดส่วนของสองประชากรที่เป็นอิสระกันในข้างต้น เนื่องจากเราไม่ทราบค่าสัดส่วนของประชากร p_1, p_2 จึงไม่สามารถหาค่าความแปรปรวนได้ ดังนั้นจึงใช้ค่าสัดส่วนตัวอย่าง $\hat{p}_1, \hat{p}_2$ ซึ่งเป็นค่าประมาณแทน จะได้ช่วงความเชื่อมั่น $(1-\alpha)100\%$ โดยประมาณ สำหรับ $p_1 - p_2$ คือ

$$(\hat{p}_1 - \hat{p}_2) \pm Z_{1-\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$$

โดยที่ $\hat{p}_i = \frac{\sum_{j=1}^{n_i} X_{ij}}{n_i}, i = 1,2$ เป็น ค่าสัดส่วนตัวอย่างของประชากรที่ 1 และ 2 ตามลำดับ ,

X_{ij} เป็นตัวอย่างสุ่มที่มีการแจกแจงแบบแบร์นูลลี และ $\hat{q}_i = 1 - \hat{p}_i, i = 1,2$

$Z_{1-\alpha/2}$ เป็นเปอร์เซ็นไทล์ (Percentile) ที่ $(1-\alpha/2)100$ ของการแจกแจงปกติมาตรฐาน

2. <u>วิธีการประมาณของนิวคอมบ์</u> (The Newcombe 's Confidence Interval)

วิธีการของ Newcombe ได้แนวความคิดมาจากวิธีการประมาณแบบช่วงด้วยรากของสมการกำลังสองสำหรับค่าสัดส่วนประชากร โดยในกรณีของการประมาณค่าแบบช่วงผลต่างค่าสัดส่วนของสองประชากร Newcombe (1998) ได้เสนอให้ใช้วิธีการประมาณที่ขั้นตอนดังนี้

กำหนดให้ (l_i, u_i) เป็นช่วงความเชื่อมั่นของ p_i ของประชากรที่ i ; i = 1, 2 เท่ากับรากของ p_i ในสมการกำลังสอง (Quadratic equation)

1) สำหรับ (l_1, u_1) ที่เป็นรากของสมการกำลังสองของ p_1 หาได้จาก

$$\frac{|\hat{p}_1 - p_1|}{\{p_1(1-p_1)/n_1\}^{1/2}} \le Z_{1-\alpha/2}$$

นำมายกกำลังสองทั้งสองข้างจะได้

$$(\hat{p}_1 - p_1)^2 - (Z_{1-\alpha/2})^2 p_1(1-p_1)/n_1 \le 0$$

กำหนดให้สมการมีค่าเท่ากับ 0 จะได้

$$\hat{p}_1^2 - 2\hat{p}_1 p_1 + p_1^2 - Z^2\frac{p_1}{n_1} + Z^2\frac{p_1^2}{n_1} = 0 \quad , Z = Z_{1-\alpha/2}$$

$$\hat{p}_1^2 - p_1\left[2\hat{p}_1 + \frac{Z^2}{n_1}\right] + p_1^2\left[1 + \frac{Z^2}{n_1}\right] = 0 \qquad \text{..........(2.1)}$$

จะเห็นได้ว่าอยู่ในรูปสมการกำลังสอง (Quadratic equation)

$$ax^2 + bx + C = 0$$

จะได้รากของสมการกำลังสองคือ

$$x = \frac{-b \pm \left(b^2 - 4ac\right)^{1/2}}{2a}$$

ดังนั้นจากสมการ (2.1) เมื่อเทียบกับรูปสมการกำลังสอง จะได้

$$x = p_1, a = 1 + \frac{Z^2}{n_1}, b = -\left(2\hat{p}_1 + \frac{Z^2}{n_1}\right) \text{ และ } c = \hat{p}_1^2$$

และจะได้ว่า
$$p_1 = \frac{\left(2\hat{p}_1 + \frac{Z^2}{n_1}\right) \pm \sqrt{\left(2\hat{p}_1 + \frac{Z^2}{n_1}\right)^2 - 4\left(1 + \frac{Z^2}{n_1}\right)\hat{p}_1^2}}{2\left(1 + \frac{Z^2}{n_1}\right)}$$

$$p_1 = \frac{\hat{p}_1 + \frac{Z^2}{2n_1} \pm \frac{Z}{2\sqrt{n_1}}\sqrt{4\hat{p}_1(1-\hat{p}_1) + \frac{Z^2}{n_1}}}{1 + \frac{Z^2}{n_1}}$$

$$p_1 = \frac{\hat{p}_1 + \frac{Z^2}{2n_1} \pm Z\sqrt{\hat{p}_1(1-\hat{p}_1)/n_1 + Z^2/4n_1^2}}{1 + \frac{Z^2}{n_1}}$$

นั่นคือช่วงความเชื่อมั่น $(1-\alpha)100\%$ สำหรับ p_1 : (l_1, u_1) คือ

$$\left(\frac{\hat{p}_1 + \frac{Z_{1-\alpha/2}^2}{2n_1} - Z_{1-\alpha/2}\sqrt{\hat{p}_1(1-\hat{p}_1)/n_1 + Z_{1-\alpha/2}^2/4n_1^2}}{1 + \frac{Z_{1-\alpha/2}^2}{n_1}}, \frac{\hat{p}_1 + \frac{Z_{1-\alpha/2}^2}{2n_1} + Z_{1-\alpha/2}\sqrt{\hat{p}_1(1-\hat{p}_1)/n_1 + Z_{1-\alpha/2}^2/4n_1^2}}{1 + \frac{Z_{1-\alpha/2}^2}{n_1}}\right)$$

2) สำหรับ (l_2, u_2) ที่เป็นรากของสมการกำลังสองของ p_2 หาได้จากวิธีการเช่นเดียวกัน จะได้ช่วงความเชื่อมั่น $(1-\alpha)100\%$ สำหรับ p_2 : (l_2, u_2) คือ

$$\left(\frac{\hat{p}_2 + \frac{Z_{1-\alpha/2}^2}{2n_2} - Z_{1-\alpha/2}\sqrt{\hat{p}_2(1-\hat{p}_2)/n_2 + Z_{1-\alpha/2}^2/4n_2^2}}{1 + \frac{Z_{1-\alpha/2}^2}{n_2}}, \frac{\hat{p}_2 + \frac{Z_{1-\alpha/2}^2}{2n_2} + Z_{1-\alpha/2}\sqrt{\hat{p}_2(1-\hat{p}_2)/n_2 + Z_{1-\alpha/2}^2/4n_2^2}}{1 + \frac{Z_{1-\alpha/2}^2}{n_2}}\right)$$

เมื่อเราทราบช่วงความเชื่อมั่น (l_i, u_i) ของค่า p_i , $i = 1,2$ ของสองประชากรแล้ว จะนำค่า (l_i, u_i) ที่ได้มาใช้ในการหาช่วงความเชื่อมั่นของผลต่างค่าสัดส่วนของสองประชากรดังนี้

กำหนดให้ $L = \hat{\theta} - \delta$ ....(2.2)

$U = \hat{\theta} - \varepsilon$ ....(2.3)

โดยที่ (L,U) เท่ากับช่วงความเชื่อมั่นของผลต่างค่าสัดส่วนของสองประชากร ($p_1 - p_2$)

$\hat{\theta}$ เท่ากับผลต่างค่าสัดส่วนตัวอย่างของสองประชากร $(\hat{p}_1 - \hat{p}_2)$

δ เท่ากับระยะทางระหว่างจุด (l_1, u_2) และจุด $(\hat{p}_1, \hat{p}_2)$

จะได้ว่า $\delta = \sqrt{(\hat{p}_1 - l_1)^2 + (u_2 - \hat{p}_2)^2}$ ....(2.4)

พิจารณาจาก $\dfrac{|\hat{p}_1 - p_1|}{\sqrt{p_1(1-p_1)/n_1}} = Z$ ....(2.5)

แทนค่า $p_1 = l_1$ ในสมการ (2.5) จะได้

$$\frac{|\hat{p}_1 - l_1|}{\sqrt{l_1(1-l_1)/n_1}} = Z$$

$$|\hat{p}_1 - l_1| = Z\sqrt{l_1(1-l_1)/n_1}$$

ยกกำลังสองทั้งสองข้างจะได้ $(\hat{p}_1 - l_1)^2 = Z^2[l_1(1-l_1)/n_1]$ ....(2.6)

และจาก $\dfrac{|\hat{p}_2 - p_2|}{\sqrt{p_2(1-p_2)/n_2}} = Z$ ....(2.7)

แทนค่า $p_2 = u_2$ ในสมการ (2.7) จะได้

$$\frac{|u_2 - \hat{p}_2|}{\sqrt{u_2(1-u_2)/n_2}} = Z$$

$$|u_2 - \hat{p}_2| = Z\sqrt{u_2(1-u_2)/n_2}$$

ยกกำลังสองทั้งสองข้างจะได้ $(u_2 - \hat{p}_2)^2 = Z^2[u_2(1-u_2)/n_2]$ ....(2.8)

แทนค่าที่ได้จากสมการที่ (2.6) และ (2.8) ลงในสมการที่ (2.4) จะได้

$$\delta = Z\sqrt{l_1(1-l_1)/n_1 + u_2(1-u_2)/n_2} \quad \text{.........(2.9)}$$

กำหนดให้ ε เท่ากับระยะทางระหว่างจุด (l_2, u_1) และจุด $(\hat{p}_1, \hat{p}_2)$

จะได้ว่า $\varepsilon = \sqrt{(u_1 - \hat{p}_1)^2 + (\hat{p}_2 - l_2)^2}$

ทำการคำนวณในลักษณะเดียวกันกับค่า δ จะได้

$$\varepsilon = Z\sqrt{u_1(1-u_1)/n_1 + l_2(1-l_2)/n_2} \quad \text{.........(2.10)}$$

นำค่า δ ที่ได้จากสมการ (2.9) แทนในสมการที่ (2.2) จะได้

$$L = \hat{\theta} - Z\sqrt{l_1(1-l_1)/n_1 + u_2(1-u_2)/n_2}$$

และค่า ε ที่ได้จากสมการที่ (2.10) แทนในสมการที่ (2.3) จะได้

$$U = \hat{\theta} + Z\sqrt{u_1(1-u_1)/n_1 + l_2(1-l_2)/n_2}$$

โดยที่ $\hat{\theta} = \hat{p}_1 - \hat{p}_2$ และ $Z = Z_{1-\alpha/2}$

ดังนั้นช่วงความเชื่อมั่นสำหรับผลต่างของค่าสัดส่วนของ 2 ประชากร (L, U) จะได้ดังนี้

$$\left(\hat{p}_1 - \hat{p}_2 - Z_{1-\alpha/2}\sqrt{\frac{l_1(1-l_1)}{n_1} + \frac{u_2(1-u_2)}{n_2}}, \hat{p}_1 - \hat{p}_2 + Z_{1-\alpha/2}\sqrt{\frac{u_1(1-u_1)}{n_1} + \frac{l_2(1-l_2)}{n_2}}\right)$$

โดยที่ค่า (l_1, u_1) เท่ากับ

$$\left(\frac{\hat{p}_1 + \frac{Z^2_{1-\alpha/2}}{2n_1} - Z_{1-\alpha/2}\sqrt{\hat{p}_1(1-\hat{p}_1)/n_1 + Z^2_{1-\alpha/2}/4n_1^2}}{1 + \frac{Z^2_{1-\alpha/2}}{n_1}}, \frac{\hat{p}_1 + \frac{Z^2_{1-\alpha/2}}{2n_1} + Z_{1-\alpha/2}\sqrt{\hat{p}_1(1-\hat{p}_1)/n_1 + Z^2_{1-\alpha/2}/4n_1^2}}{1 + \frac{Z^2_{1-\alpha/2}}{n_1}}\right)$$

และ (l_2, u_2) เท่ากับ

$$\left(\frac{\hat{p}_2+\frac{Z_{1-\alpha/2}^2}{2n_2}-Z_{1-\alpha/2}\sqrt{\hat{p}_2(1-\hat{p}_2)/n_2+Z_{1-\alpha/2}^2/4n_2^2}}{1+\frac{Z_{1-\alpha/2}^2}{n_2}},\frac{\hat{p}_2+\frac{Z_{1-\alpha/2}^2}{2n_2}+Z_{1-\alpha/2}\sqrt{\hat{p}_2(1-\hat{p}_2)/n_2+Z_{1-\alpha/2}^2/4n_2^2}}{1+\frac{Z_{1-\alpha/2}^2}{n_2}}\right)$$

3. <u>วิธีการประมาณของเจฟฟรีย์</u> (The Jeffrey 's Confidence Interval)

วิธีการนี้เป็นวิธีการที่นำเอาแนวความคิดของเบส์มาประยุกต์ใช้ในการประมาณค่าสัดส่วนประชากรโดยกำหนดให้การแจกแจงแรกเริ่ม (Prior Distribution) มีการแจกแจงแบบเบตา ซึ่งจะได้การแจกแจงภายหลัง (Posterior Distribution) เพื่อใช้ในการหาตัวประมาณเบส์ ที่ใช้ในการประมาณค่าสัดส่วนประชากร

กำหนดให้ $X_1, X_2, \ldots\ldots, X_n$ เป็นตัวอย่างสุ่มที่มีการแจกแจงแบบแบร์นูลลี โดยมีพารามิเตอร์ p เขียนแทนด้วย $X \sim Ber(p)$ และ $Y=\sum_{i=1}^{n} X_i$ มีการแจกแจงแบบทวินามโดยมีพารามิเตอร์ n และ p เขียนแทนด้วย $Y \sim B(n,p)$

ฟังก์ชันความหนาแน่นมีเงื่อนไขของ Y เมื่อกำหนด p คือ

$$f(y/p)=\binom{n}{y}p^y(1-p)^{n-y} \ ,\ 0<p<1$$

p มีฟังก์ชันความหนาแน่นแรกเริ่ม (Prior density function) $Beta(a_1,a_2)$ คือ

$$g(p) = \frac{\Gamma(a_1+a_2)}{\Gamma(a_1)\Gamma(a_2)}p^{a_1-1}(1-p)^{a_2-1}, 0<p<1$$

และฟังก์ชันความหนาแน่นภายหลัง (Posterior density function) ของ p หาได้ดังนี้

$$g(p/y) = \frac{g(p)f(y/p)}{\int_0^1 g(p)f(y/p)dp}$$

$$\int_0^1 g(p)f(y/p)dp = \int_0^1 \binom{n}{y}\frac{\Gamma(a_1+a_2)}{\Gamma(a_1)\Gamma(a_2)}p^{a_1+y-1}(1-p)^{a_2+n-y+1}dp$$

$$= \binom{n}{y}\frac{\Gamma(a_1+a_2)\Gamma(a_1')\Gamma(a_2')}{\Gamma(a_1)\Gamma(a_2)\Gamma(a_1'+a_2')}$$

ซึ่ง $a_1' = y + a_1, a_2' = n - y + a_2$

นั่นคือ $$g(p/y) = \frac{\Gamma(a_1' + a_2')}{\Gamma(a_1')\Gamma(a_2')} p^{y+a_1-1}(1-p)^{n-y+a_2-1}$$

$$= \frac{\Gamma(a_1' + a_2')}{\Gamma(a_1')\Gamma(a_2')} p^{a_1'-1}(1-p)^{a_2'-1}$$

ดังนั้นฟังก์ชันความหนาแน่นภายหลังของ p คือ $Beta(a_1', a_2')$

กำหนดให้ $l(p,\delta)$ แทนฟังก์ชันความสูญเสียเป็นฟังก์ชันกำลังสองนั่นคือ

$$l(p,\delta) = (p-\delta)^2$$

ดังนั้นตัวประมาณเบส์ $(\delta_B(Y))$คือ $$\hat{p} = E(p/Y) = \frac{a_1'}{a_1' + a_2'}$$

$$= \frac{Y + a_1}{n + a_1 + a_2}$$

Brown, Cai and DasGupta (2001) เสนอว่าควรจะให้การแจกแจงแรกเริ่มเป็น $Beta(1/2, 1/2)$ และจากข้อมูลข้างต้นเมื่อแทนค่า $a_1 = 1/2$ และ $a_2 = 1/2$ จะได้ตัวประมาณเบส์ในกรณีการประมาณค่าแบบช่วงสำหรับผลต่างของค่าสัดส่วนของ 2 ประชากรดังนี้

$$\tilde{p}_1 = \frac{(Y_1 + 1/2)}{(n_1 + 1)} \qquad \tilde{p}_2 = \frac{(Y_2 + 1/2)}{(n_2 + 1)}$$

โดยที่ Y_1 เป็นจำนวนครั้งของการเกิดลักษณะที่สนใจจากการทดลองแบร์นูลลี n_1 ครั้งในประชากรที่ 1 $= \sum_{j=1}^{n_1} X_{1j}$, X_{1j} เป็นตัวอย่างสุ่มที่มีการแจกแจงแบบแบร์นูลลีในประชากรที่ 1

Y_2 เป็นจำนวนครั้งของการเกิดลักษณะที่สนใจจากการทดลองแบร์นูลลี n_2 ครั้งในประชากรที่ 2 $= \sum_{j=1}^{n_2} X_{2j}$, X_{2j} เป็นตัวอย่างสุ่มที่มีการแจกแจงแบบแบร์นูลลีในประชากรที่ 2

นำตัวประมาณเบส์ที่ได้ $\tilde{p}_1, \tilde{p}_2$ ไปแทนที่ค่า $\hat{p}_1, \hat{p}_2$ ในช่วงความเชื่อมั่น $(1-\alpha)100\%$ สำหรับ p_1, p_2 ที่ได้จากวิธีประมาณแบบฉบับจะได้ช่วงความเชื่อมั่น $(1-\alpha)100\%$ สำหรับ $p_1 - p_2$ คือ

$$\left(\tilde{p}_1 - \tilde{p}_2 \pm Z_{1-\alpha/2}\sqrt{\frac{\tilde{p}_1\tilde{q}_1}{n_1} + \frac{\tilde{p}_2\tilde{q}_2}{n_2}}\right) \quad ; \tilde{q}_1 = 1 - \tilde{p}_1, \tilde{q}_2 = 1 - \tilde{p}_2$$

โดยที่ $\tilde{p}_1 = \dfrac{(Y_1 + 1/2)}{(n_1 + 1)}$, $\tilde{q}_1 = 1 - \tilde{p}_1$

$$\tilde{p}_2 = \frac{(Y_2 + 1/2)}{(n_2 + 1)},\ \tilde{q}_2 = 1 - \tilde{p}_2$$

4. <u>วิธีการประมาณแบบจัดค่ากลางใหม่</u> (The Recentered Confidence Interval)

การประมาณค่าแบบช่วงสำหรับผลต่างระหว่างค่าสัดส่วนของสองประชากร Lawrence, Brown และ Xuefeng Li(2003) ได้เสนอให้ทำการสร้างค่าพารามิเตอร์ใหม่ (Reparametrazation) แล้วสร้างช่วงความเชื่อมั่นสำหรับผลต่างระหว่างค่าสัดส่วนของสองประชากร โดยมีขั้นตอนดังนี้

กำหนดให้ Δ เท่ากับผลต่างระหว่างค่าสัดส่วนของสองประชากร $= p_1 - p_2 \geq 0$, p_1, p_2 เป็นค่าสัดส่วนประชากรในประชากรที่1 และ 2 ตามลำดับ $(0 \leq p_1, p_2 \leq 1)$ และพารามิเตอร์ที่สร้างขึ้นใหม่ คือ

$$p = \frac{(p_1/n_1) + (p_2/n_2)}{(1/n_1) + (1/n_2)} = \frac{(n_2 p_1 + n_1 p_2)}{n_1 + n_2} \qquad \text{..........(2.11)}$$

เมื่อแทนค่า $p_2 = p_1 - \Delta$ ในสมการ (2.11) จะได้

$$p = \frac{n_2 p_1 + n_1 (p_1 - \Delta)}{n_1 + n_2}$$

$$(n_1 + n_2)p = n_2 p_1 + n_1 p_1 - \Delta n_1$$

$$(n_1 + n_2)p + \Delta n_1 = p_1 (n_1 + n_2)$$

$$p_1 = \frac{(n_1 + n_2)p + \Delta n_1}{n_1 + n_2}$$

$$p_1 = p + \frac{\Delta n_1}{n_1 + n_2} \qquad \text{..........(2.12)}$$

และแทนค่า $p_1 = \Delta + p_2$ ในสมการ (2.11) เช่นกันจะได้

$$p_2 = p - \frac{\Delta n_2}{n_1 + n_2} \qquad \text{..........(2.13)}$$

การพิจารณาว่าค่า p ตามสมการ (2.11) มีค่าอยู่ในช่วงใด ทำได้ดังนี้

จากข้อกำหนดที่ว่า $\Delta = p_1 - p_2 \geq 0$ $(p_1 \geq p_2)$ และเราทราบว่าค่า p_1, p_2 มีค่าสูงสุดเท่ากับ 1 และมีค่าต่ำสุดเท่ากับ 0 $(0 \leq p_1, p_2 \leq 1)$ ดังนั้นจะได้ว่า

- ค่าสูงสุดของค่า p ได้มาจากการแทนค่า $p_1 = 1$ ในสมการ (2.12) จะได้

  p มีค่าสูงสุดเท่ากับ $1 - \dfrac{\Delta n_1}{n_1 + n_2}$

- ค่าต่ำสุดของค่า p ได้มาจากการแทนค่า $p_2 = 0$ ในสมการ (2.13) จะได้

  p มีค่าต่ำสุดเท่ากับ $\dfrac{\Delta n_2}{n_1 + n_2}$

สรุปได้ว่า ค่า p มีค่าอยู่ในช่วง $\left(\dfrac{\Delta n_2}{n_1 + n_2}, 1 - \dfrac{\Delta n_1}{n_1 + n_2}\right)$

ในการหาค่าความแปรปรวนของผลต่างระหว่างค่าสัดส่วนของสองประชากรที่ได้จากค่าพารามิเตอร์ใหม่ (p) สามารถทำได้โดยพิจารณาจากสมการค่าความแปรปรวนของผลต่างระหว่างค่าสัดส่วนของสองประชากรที่ได้จากค่าพารามิเตอร์เดิม (p_1, p_2) ทำได้โดย

จาก

$$
\begin{aligned}
\sigma_{\Delta}^2 &= \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2} \\
&= \frac{p_1 - p_1^2}{n_1} + \frac{p_2 - p_2^2}{n_2} \\
&= \frac{p_1(n_1+n_2)}{n_1(n_1+n_2)} - \frac{p_1^2}{n_1} + \frac{p_2(n_1+n_2)}{n_2(n_1+n_2)} - \frac{p_2^2}{n_2} \\
&= \frac{n_1p_1 + n_2p_1}{n_1(n_1+n_2)} - \frac{p_1^2(n_1+n_2)}{n_1(n_1+n_2)} + \frac{n_1p_2 + n_2p_2}{n_2(n_1+n_2)} - \frac{p_2^2(n_1+n_2)}{n_2(n_1+n_2)} \\
&= \frac{n_2p_1 - n_2p_1^2 - n_1p_1^2}{n_1(n_1+n_2)} + \frac{p_1}{(n_1+n_2)} + \frac{n_1p_2 - n_1p_2^2 - n_2p_2^2}{n_2(n_1+n_2)} \\
&\quad + \frac{p_2}{(n_1+n_2)} + \frac{2p_1p_2 - 2p_1p_2}{n_1+n_2} \\
&= \frac{n_2p_1(1-p_1)}{n_1(n_1+n_2)} - \frac{p_1^2}{(n_1+n_2)} + \frac{n_1p_2(1-p_2)}{n_2(n_1+n_2)} - \frac{p_2^2}{(n_1+n_2)} \\
&\quad + \frac{p_1 - p_1p_2}{(n_1+n_2)} + \frac{2p_1p_2}{n_1+n_2}
\end{aligned}
$$

$$= \frac{n_2 p_1 q_1}{n_1(n_1+n_2)} + \frac{n_1 p_2 q_2}{n_2(n_1+n_2)} + \frac{p_1 q_2}{(n_1+n_2)} + \frac{p_2 q_1}{(n_1+n_2)} - \frac{(p_1-p_2)^2}{(n_1+n_2)}$$

$$= \frac{n_2 p_1 q_1 + n_1 p_1 q_2}{n_1(n_1+n_2)} + \frac{n_1 p_2 q_2 + n_2 p_2 q_1}{n_2(n_1+n_2)} - \frac{(p_1-p_2)^2}{(n_1+n_2)}$$

$$= \frac{p_1}{n_1}\left[\frac{n_2 q_1 + n_1 q_2}{n_1+n_2}\right] + \frac{p_2}{n_2}\left[\frac{n_1 q_2 + n_2 q_1}{n_1+n_2}\right] - \frac{\Delta^2}{n_1+n_2}$$

$$= \left[\frac{p_1}{n_1} + \frac{p_2}{n_2}\right]\left[\frac{n_1 q_2 + n_2 q_1}{n_1+n_2}\right] - \frac{\Delta^2}{n_1+n_2}$$

$$= \left[\frac{n_2 p_1 + n_1 p_2}{n_1 n_2}\frac{(n_1+n_2)}{(n_1+n_2)}\right](q) - \frac{\Delta^2}{n_1+n_2}$$

$$= \left(\frac{n_1+n_2}{n_1 n_2}\right)\left(\frac{n_2 p_1 + n_1 p_2}{n_1+n_2}\right)(q) - \frac{\Delta^2}{n_1+n_2}$$

จะได้ $\sigma^2_\Delta \quad = \left(\frac{1}{n_1} + \frac{1}{n_2}\right)pq - \frac{\Delta^2}{n_1+n_2}$ ....(2.14)

โดยที่ p มีค่าเท่ากับสมการ (2.11) , $q = 1 - p$ และ $\Delta = p_1 - p_2$

เนื่องจากเราไม่สามารถทราบค่า p_1, p_2 ได้ จึงใช้ค่าประมาณแทน $(\hat{p}_1, \hat{p}_2)$ ดังนั้นจากสมการที่ (2.11) จะได้ค่าประมาณของ p $(\hat{p})$ คือ

$$\hat{p} \quad = \frac{(n_2 \hat{p}_1 + n_1 \hat{p}_2)}{n_1 + n_2}$$

โดยที่ $\hat{p}_i = \frac{\sum_{j=1}^{n_i} X_{ij}}{n_i}, i = 1,2$ เป็น ค่าสัดส่วนตัวอย่างของประชากรที่ 1 และ 2 ตามลำดับ , X_{ij} เป็นตัวอย่างสุ่มที่มีการแจกแจงแบบแบร์นูลลี และ $\hat{q}_i = 1 - \hat{p}_i, i = 1,2$

และค่า $\hat{p}$ ที่ได้นี้ก็มีค่าอยู่ในช่วงเช่นเดียวกับค่า p กล่าวคือ ค่า $\hat{p}$ มีค่าอยู่ในช่วง $\left(\frac{\Delta n_2}{n_1+n_2}, 1 - \frac{\Delta n_1}{n_1+n_2}\right)$ หรือกำหนดให้เป็นค่าประมาณตัวใหม่ที่มีเงื่อนไขดังนี้

$$\begin{cases} \Delta n_2/(n_1+n_2) & , \hat{p} < \Delta n_2/(n_1+n_2) \\ & \\ & \end{cases}$$

$$\widetilde{p} = \begin{cases} \hat{p} & , \Delta n_2/(n_1+n_2) \le \hat{p} \le 1-\Delta n_1/(n_1+n_2) \\ 1-\Delta n_1/(n_1+n_2) & , \hat{p} > \Delta n_1/(n_1+n_2) \end{cases} \quad \text{.........(2.15)}$$

เนื่องจากไม่ทราบค่า σ_Δ^2 จึงประมาณด้วย S_Δ^2 ดังนั้นสมการ (2.14) จะได้

$$S_\Delta^2 = \left(\frac{1}{n_1}+\frac{1}{n_2}\right)\hat{p}\hat{q} - \frac{\hat{\Delta}^2}{n_1+n_2}$$

แทนค่า $\hat{p} = \widetilde{p}$ ที่ได้จากสมการ (2.15) จะได้

$$S_\Delta^2 = \left(\frac{1}{n_1}+\frac{1}{n_2}\right)\widetilde{p}\widetilde{q} - \frac{\hat{\Delta}^2}{n_1+n_2} \quad \text{..........(2.16)}$$

ในการหาช่วงความเชื่อมั่นสำหรับผลต่างระหว่างค่าสัดส่วนของสองประชากรในกรณีที่เราไม่ทราบค่า σ_Δ^2 พิจารณาจากตัวสถิติดังนี้

$$t = \frac{\hat{\Delta}-\Delta}{\sqrt{S_\Delta^2}}$$ มีการแจกแจงแบบทีด้วยองศาอิสระ n_1+n_2-2

โดยที่ $\Delta = p_1 - p_2$, $\hat{\Delta} = \hat{p}_1 - \hat{p}_2$

พิจารณาจาก

$$\frac{\hat{\Delta}-\Delta}{\sqrt{S_\Delta^2}} \le t_{1-\alpha/2}$$

$$\Delta - \hat{\Delta} \le t_{1-\alpha/2}\sqrt{S_\Delta^2}$$

นำมายกกำลังสองทั้งสองข้างจะได้

$$\left(\hat{\Delta}-\Delta\right)^2 - t_{1-\alpha/2}{}^2 S_\Delta^2 \le 0$$

กำหนดให้สมการมีค่าเท่ากับ 0 และแทนค่า S_Δ^2 ที่ได้จากสมการ (2.16) จะได้

$$\left(\hat{\Delta}-\Delta\right)^2 - t^2\left[\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\widetilde{p}\widetilde{q} - \frac{\hat{\Delta}^2}{n_1+n_2}\right] = 0 \text{ , } t = t_{1-\alpha/2}$$

$$\hat{\Delta}^2 - 2\hat{\Delta}\Delta + \Delta^2 - t^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\widetilde{p}\widetilde{q} + t^2\,\frac{\Delta^2}{n_1+n_2} = 0$$

$$\left(1+\frac{t^2}{n_1+n_2}\right)\Delta^2 - 2\hat{\Delta}\Delta + \hat{\Delta}^2 - t^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\widetilde{p}\widetilde{q} \qquad = 0$$

จะเห็นได้ว่าอยู่ในรูปสมการกำลังสอง (Quadratic equation)

$$ax^2 + bx + C = 0$$

จะได้รากของสมการกำลังสองคือ

$$x = \frac{-b \pm \left(b^2 - 4ac\right)^{1/2}}{2a}$$

เมื่อ $x = \Delta, a = 1 + \dfrac{t^2}{n_1+n_2}, b = -2\hat{\Delta}$ และ $c = \hat{\Delta}^2 - t^2\left(\dfrac{1}{n_1}+\dfrac{1}{n_2}\right)\widetilde{p}\widetilde{q}$

ดังนั้น

$$\Delta = \frac{2\hat{\Delta} \pm \sqrt{4\hat{\Delta}^2 - 4\left(1-\dfrac{t^2}{n_1+n_2}\right)\left[\hat{\Delta}^2 - t^2\left(\dfrac{1}{n_1}+\dfrac{1}{n_2}\right)\widetilde{p}\widetilde{q}\right]}}{2\left(1+\dfrac{t^2}{n_1+n_2}\right)}$$

$$\Delta = \frac{\hat{\Delta}}{1+\dfrac{t^2}{n_1+n_2}} \pm \frac{\sqrt{\hat{\Delta}^2 - \left(1-\dfrac{t^2}{n_1+n_2}\right)\left[\hat{\Delta}^2 - t^2\left(\dfrac{1}{n_1}+\dfrac{1}{n_2}\right)\widetilde{p}\widetilde{q}\right]}}{\left(1+\dfrac{t^2}{n_1+n_2}\right)}$$

$$\Delta = \frac{\hat{\Delta}}{1+\dfrac{t^2}{n_1+n_2}} \pm \frac{\sqrt{\hat{\Delta}^2 - \hat{\Delta}^2 + t^2\widetilde{p}\widetilde{q}\left(\dfrac{1}{n_1}+\dfrac{1}{n_2}\right) + \dfrac{t^2\widetilde{p}\widetilde{q}}{n_1+n_2}\left(\dfrac{1}{n_1}+\dfrac{1}{n_2}\right) - \dfrac{t^2\hat{\Delta}^2}{n_1+n_2}}}{\left(1+\dfrac{t^2}{n_1+n_2}\right)}$$

$$\Delta = \frac{\hat{\Delta}}{1+\dfrac{t^2}{n_1+n_2}} \pm \frac{\sqrt{t^2\left[\left(\dfrac{1}{n_1}+\dfrac{1}{n_2}\right)\widetilde{p}\widetilde{q} + \dfrac{t^2\widetilde{p}\widetilde{q}}{n_1+n_2}\left(\dfrac{1}{n_1}+\dfrac{1}{n_2}\right) - \dfrac{\hat{\Delta}^2}{n_1+n_2}\right]}}{\left(1+\dfrac{t^2}{n_1+n_2}\right)}$$

$$\Delta = \frac{\hat{\Delta}}{1+\frac{t^2}{n_1+n_2}} \pm \frac{t\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\widetilde{p}\widetilde{q}\left(1+\frac{t^2}{n_1+n_2}\right)-\frac{\hat{\Delta}^2}{n_1+n_2}}}{\left(1+\frac{t^2}{n_1+n_2}\right)}$$

ดังนั้นจะได้ช่วงความเชื่อมั่น $(1-\alpha)100\%$ สำหรับ $p_1 - p_2$ คือ

$$\left(\frac{\hat{\Delta}}{1+t^2/(n_1+n_2)} \pm \frac{t\sqrt{[1+t^2/(n_1+n_2)][(1/n_1+1/n_2)\widetilde{p}\widetilde{q}-\hat{\Delta}^2/(n_1+n_2)]}}{1+t^2/(n_1+n_2)}\right)$$

โดยที่ $\hat{\Delta} = \hat{p}_1 - \hat{p}_2$, $t = t_{1-\alpha/2}$

$$\widetilde{p} = \begin{cases} \Delta n_2/(n_1+n_2) & \text{ถ้า} \quad \hat{p} < \Delta n_2/(n_1+n_2) \\ \hat{p} & \text{ถ้า} \quad \Delta n_2/(n_1+n_2) \le \hat{p} \le 1-\Delta n_1/(n_1+n_2) \\ 1-\Delta n_1/(n_1+n_2) & \text{ถ้า} \quad \hat{p} > \Delta n_1/(n_1+n_2) \end{cases}$$

โดยเรียกวิธีการนี้ว่า วิธีการประมาณแบบจัดค่ากลางใหม่ (The Recentered Confidence Interval) เนื่องจากสูตรช่วงความเชื่อมั่นที่ได้มีการปรับค่าใหม่โดยใช้ค่า $\left(1+\frac{k^2}{(n_1+n_2)}\right)^{-1}$

2.5 เกณฑ์ที่ใช้ในการเปรียบเทียบวิธีการประมาณค่าแบบช่วงสำหรับผลต่างค่าสัดส่วนสองประชากร

การเปรียบเทียบวิธีการประมาณค่าแบบช่วงสำหรับผลต่างค่าสัดส่วนประชากร 2 กลุ่ม จะทำการเปรียบเทียบค่าสัมประสิทธิ์ความเชื่อมั่นและค่าความยาวเฉลี่ยของช่วงความเชื่อมั่นที่คำนวณได้จากแต่ละวิธี ที่ระดับความเชื่อมั่น 3 ระดับ คือ 90% , 95% และ 99% ในแต่ละสถานการณ์การทดลอง ทำการทดลองซ้ำ 1,000 ครั้ง ในการตรวจสอบว่าวิธีการประมาณใดให้ค่าสัมประสิทธิ์ความเชื่อมั่นจากการทดลองไม่ต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนดได้หรือไม่นั้น ผู้วิจัยอาศัยการทดสอบสมมติฐานโดยใช้ตัวสถิติ Z ดังนี้

$$H_0 : p \ge p_0$$

$H_1 : p < p_0$

พิจารณา ความน่าจะเป็นที่จะยอมรับ H_0 เมื่อ H_0 เป็นจริงที่ระดับนัยสำคัญ α_0 :

$$P\left(-Z_{1-\alpha_0} < \frac{\hat{p}-p_0}{\sqrt{\frac{p_0(1-p_0)}{N}}} \right) = 1-\alpha_0$$

ดังนั้น

$$-Z_{1-\alpha_0}\sqrt{\frac{p_0(1-p_0)}{N}} < \hat{p}-p_0$$

$$p_0 - Z_{1-\alpha_0}\sqrt{\frac{p_0(1-p_0)}{N}} < \hat{p}$$

เมื่อพิจารณาค่า $0 < p < 1$ จะได้ว่า

$$p_0 - Z_{1-\alpha_0}\sqrt{\frac{p_0(1-p_0)}{N}} < \hat{p} < 1$$

ดังนั้นจะได้ช่วงของการยอมรับสมมติฐานหลัก H_0 คือ

$$\left(p_0 - Z_{1-\alpha_0}\sqrt{\frac{p_0(1-p_0)}{N}}, 1 \right)$$

โดยที่ α_0 = ระดับนัยสำคัญหรือความน่าจะเป็นที่จะเกิดความผิดพลาดที่กำหนดในการทดสอบสำหรับการวิจัยครั้งนี้กำหนดระดับนัยสำคัญของการทดสอบเท่ากับ 0.05

N = จำนวนครั้งของการทดลอง (2,000)

p_0 = ค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด (0.90, 0.95 และ 0.99)

$\hat{p}$ = ค่าสัมประสิทธิ์ความเชื่อมั่นที่ได้จากการทดลอง หรือความน่าจะเป็นที่วิธีการประมาณนั้นจะคลุมค่า $p_1 - p_2$ ซึ่งหาได้จากจำนวนครั้งที่คลุมค่า $p_1 - p_2$ หารด้วยจำนวนครั้งในการทดลอง (N)

สรุปได้ว่า ถ้าวิธีการประมาณใดให้ค่าสัมประสิทธิ์ความเชื่อมั่นจากการทดลองอยู่ในช่วงที่ยอมรับสมมติฐาน จะถือว่าวิธีการประมาณนั้นให้ค่าสัมประสิทธิ์ความเชื่อมั่นจากการทดลองไม่ต่ำกว่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด

1. <u>ระดับความเชื่อมั่น 90 %</u>

$H_0 : p \geq 0.90$
$H_1 : p < 0.90$

จะได้ว่า วิธีการประมาณให้ค่าสัมประสิทธิ์ความเชื่อมั่นไม่ต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด ถ้าค่า $\hat{p}$ มีค่าอยู่ในช่วง

$$\left(0.90 - 1.645\sqrt{\frac{0.90(0.10)}{2,000}}, 1\right) = (0.8890, 1)$$

2. <u>ระดับความเชื่อมั่น 95 %</u>

$H_0 : p \geq 0.95$
$H_1 : p < 0.95$

จะได้ว่า วิธีการประมาณให้ค่าสัมประสิทธิ์ความเชื่อมั่นไม่ต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด ถ้าค่า $\hat{p}$ มีค่าอยู่ในช่วง

$$\left(0.95 - 1.645\sqrt{\frac{0.95(0.05)}{2,000}}, 1\right) = (0.9420, 1)$$

3. <u>ระดับความเชื่อมั่น 99 %</u>

$H_0 : p \geq 0.99$
$H_1 : p < 0.99$

จะได้ว่า วิธีการประมาณให้ค่าสัมประสิทธิ์ความเชื่อมั่นไม่ต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด ถ้าค่า $\hat{p}$ มีค่าอยู่ในช่วง

$$\left(0.99 - 1.645\sqrt{\frac{0.99(0.01)}{2,000}}, 1\right) = (0.9863, 1)$$

เพราะฉะนั้นการที่จะสรุปได้ว่าวิธีการประมาณใดให้ค่าสัมประสิทธิ์ความเชื่อมั่นไม่ต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนดนั้น ค่าสัมประสิทธิ์ความเชื่อมั่นที่ได้จากการทดลองจะต้องมีค่าไม่ต่ำกว่า 0.8890, 0.9420, 0.9863 ที่ระดับความเชื่อมั่น 90%, 95% และ 99% ตามลำดับ