# บทที่ 2

# ทฤษฎีที่เกี่ยวข้องกับงานวิจัย

เนื่องจากงานวิจัยนี้ได้เสนอวิธีการประมาณค่าสัดส่วนประชากรแบบช่วง 5 วิธีด้วยกัน ซึ่งแต่ละวิธีมีทฤษฎีที่เกี่ยวข้องแตกต่างกันไป ดังนั้นในบทนี้จึงขอนำเสนอรายละเอียดของทฤษฎีบทต่าง ๆ ที่เกี่ยวข้องกับงานวิจัย ดังต่อไปนี้

## 2.1 ทฤษฎีบทลิมิตเข้าสู่ส่วนกลาง

เมื่อสุ่มตัวอย่างขนาดใหญ่จากประชากรที่มีการแจกแจงแบบใด ๆ ก็ตามที่มีความแปรปรวนเป็นค่าจำกัด ค่าเฉลี่ยตัวอย่างจะมีการแจกแจงลู่เข้าสู่การแจกแจงแบบปกติ นั่นคือ

ถ้า $\bar{Y}$ คือค่าเฉลี่ยของตัวอย่างสุ่ม $Y_1, Y_2, \ldots, Y_n$ จากการแจกแจงที่มีค่าเฉลี่ย $\mu$ และมีค่าความแปรปรวน $\sigma^2 < \infty$ แล้ว เมื่อ $n \to \infty$ ได้ว่า

$$\bar{Y} \sim \text{normal}\left(\mu, \sigma^2_{\bar{Y}}\right) \quad \text{(โดยประมาณ)}$$

หรือ

$$Z = \frac{\bar{Y} - \mu}{\sigma/\sqrt{n}} \sim \text{normal}(0,1) \quad \text{(โดยประมาณ)}$$



## 2.2 การทดลองแบร์นูลลีและการแจกแจงทวินาม

การทดลองแบร์นูลลี เป็นการทดลองที่ในแต่ละครั้งของการทดลองประกอบไปด้วยผลการทดลองที่เป็นไปได้ 2 ลักษณะคือ ลักษณะที่สนใจ และ ลักษณะที่ไม่สนใจ ยกตัวอย่างเช่น การโยนเหรียญ 1 อัน 1 ครั้ง ผลจากการทดลองจะเป็นไปได้ 2 ลักษณะคือ หัว หรือ ก้อย อย่างใดอย่างหนึ่ง ดังนั้นถ้าลักษณะที่สนใจคือการที่เหรียญขึ้นหน้าหัว การที่เหรียญขึ้นหน้าก้อยก็ถูกจัดเป็นลักษณะที่ไม่สนใจ เป็นต้น

**การแจกแจงแบร์นูลลี**

ในแต่ละครั้งของการทดลองแบร์นูลลี ถ้าให้ค่าความน่าจะเป็นของการเกิดลักษณะที่สนใจมีค่าเท่ากับ $p$ และค่าความน่าจะเป็นของการเกิดลักษณะที่ไม่สนใจมีค่าเท่ากับ $1-p$

กำหนดให้ $X$ คือตัวแปรสุ่มที่มีค่าเป็นไปได้คือ

$$X = \begin{cases} 1 & \text{เมื่อเกิดลักษณะที่สนใจ ด้วยความน่าจะเป็น } p \\ 0 & \text{เมื่อเกิดลักษณะที่ไม่สนใจ ด้วยความน่าจะเป็น } 1-p \end{cases}$$

ฟังก์ชันความน่าจะเป็นของ $X$ เขียนได้ดังนี้

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

ค่าคาดหวังของ $X$ (Expected value of $X$) คือ

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

ค่าความแปรปรวนของ $X$ (Variance of $X$) คือ

$$\sigma^2 = \mathrm{Var(X)} = \sum_{x=0}^{1}(\mathrm{x}-\mathrm{p})^2\mathrm{p}^{x}(1-p)^{1-x} = (0-\mathrm{p})^2(1-p)+(1-p)^2(p) = \mathrm{pq}$$

**การแจกแจงทวินาม**

เมื่อทำการทดลองแบร์นูลลี $n$ ครั้งติดต่อกัน เราสนใจจำนวนครั้งรวมของการเกิดลักษณะที่สนใจโดยไม่คำนึงถึงตำแหน่งของการเกิด กำหนดให้ตัวแปรสุ่ม $Y$ คือจำนวนครั้งของการเกิดลักษณะที่สนใจจากการทดลองแบร์นูลลี $n$ ครั้ง ดังนั้นค่าที่เป็นไปได้ของ $Y$ คือ 0, 1, 2, ..., $n$ ถ้าจากการทดลองเกิดลักษณะที่สนใจ $y$ ครั้งโดยที่ $y = 0,1,\ldots,n$ และเกิดลักษณะที่ไม่สนใจ $n-y$ ครั้ง

ดังนั้นความน่าจะเป็นของการเกิดลักษณะที่สนใจ $y$ ครั้งจากการทดลอง $n$ ครั้งจึงมีค่าเท่ากับ $p^y(1-p)^{n-y}$ โดยใช้ความเป็นอิสระของการทดลองแบร์นูลลี นอกจากนั้นจำนวนหนทางของการเกิดลักษณะที่สนใจ $y$ ครั้งจากการทดลองแบร์นูลลี $n$ ครั้งมีค่าเท่ากับ

$$\binom{n}{y} = \frac{\mathrm{n!}}{\mathrm{y!(n-y)!}}$$

ดังนั้นจะได้ฟังก์ชันความน่าจะเป็นของ $Y$ คือ

$$f(y) = \binom{\mathrm{n}}{\mathrm{y}}\mathrm{p}^y(1-p)^{n-y} \qquad ;\mathrm{y} = 0,1,\ldots,\mathrm{n}\,;0 \le \mathrm{p} \le 1$$

โดยมี $n$ และ $p$ เป็นพารามิเตอร์ มีค่าคาดหวัง $\mu = \mathrm{E(Y)} = \mathrm{np}$ และค่าความแปรปรวน $\sigma^2 = \mathrm{Var(Y)} = np(1-p) = npq$

**การประมาณค่าสัดส่วนทวินาม**

เมื่อทำการทดลองแบร์นูลลี $n$ ครั้ง โดยมีค่าความน่าจะเป็นของการเกิดลักษณะที่สนใจเท่ากับ $p$ ถ้าจำนวนรวมของการเกิดลักษณะที่สนใจคือ $Y$ เมื่อต้องการหาค่าประมาณแบบจุดของค่าสัดส่วนประชากร จะได้ว่าค่าสัดส่วนตัวอย่าง

$$\hat{p} = \frac{Y}{n}$$

คือตัวประมาณสำหรับค่าสัดส่วนประชากร $p$

สำหรับคุณสมบัติของตัวประมาณ เนื่องจาก $Y$ มีการแจกแจงทวินามมีพารามิเตอร์คือ $n$ และ $p$ มีค่าเฉลี่ย $np$ และค่าความแปรปรวน $npq$ ดังนั้นเมื่อพิจารณาค่าเฉลี่ยของตัวประมาณจะได้ว่า

$$E\left(\hat{p}\right) = \frac{1}{n}E(Y) = \frac{np}{n} = p$$

แสดงให้เห็นว่า $\hat{p}$ เป็นตัวประมาณที่ไม่เอนเอียง (unbiased estimator) ของ $p$ และเมื่อพิจารณาค่าความแปรปรวนของตัวประมาณ $\hat{p}$ จะได้ว่า

$$Var\left(\hat{p}\right) = \frac{1}{n^2}Var(Y) = \frac{npq}{n^2} = \frac{\mathrm{pq}}{\mathrm{n}}$$

และตัวประมาณของความแปรปรวนจะได้จากค่าประมาณตัวอย่าง $\hat{p}$ ของ $p$ นั่นคือ

$$\hat{Var}\left(\hat{p}\right) = \frac{\hat{\mathrm{p}}\hat{\mathrm{q}}}{\mathrm{n}}$$

ถ้า $n$ มีขนาดใหญ่ จากทฤษฎีบทลิมิตเข้าสู่ส่วนกลาง $\hat{p} = Y/n$ สามารถประมาณได้ด้วยการแจกแจงแบบปกติที่มีค่าเฉลี่ย $p$ และความแปรปรวน $pq/n$ นั่นก็คือเมื่อ $n \to \infty$

$$\hat{p} \sim \mathrm{N}\left(\mathrm{p}, \frac{\mathrm{pq}}{\mathrm{n}}\right) \qquad \text{(โดยประมาณ)}$$

และ

$$Z = \frac{\hat{p} - p}{\sqrt{pq/n}} \sim \mathrm{N}(0,1) \qquad \text{(โดยประมาณ)}$$

เมื่อพิจารณาค่าความแปรปรวนของตัวแปรสุ่มทวินามพบว่า $Var\left(\hat{p}\right)$ มีค่าขึ้นอยู่กับ $E\left(\hat{p}\right)$ นั่นก็คือ

$$\sigma^2_{\hat{p}} \propto \mathrm{E}\left(\hat{\mathrm{p}}\right)\left\{1 - E\left(\hat{p}\right)\right\}$$

ซึ่งจะพบว่า ถ้าค่าประมาณ $E\left(\hat{p}\right)$ ไม่ถูกต้องจะส่งผลให้ค่าประมาณ $Var\left(\hat{p}\right)$ ไม่ถูกต้องด้วย และจากลักษณะดังกล่าว Montgomery and Peck (1982) แนะนำให้ทำการแปลงเชิงเส้นค่า $\hat{p}$ ให้อยู่ในรูปของอาร์คไซน์ ซึ่งจะมีผลให้ค่าประมาณ $Var\left(\hat{p}\right)$ เป็นค่าคงที่ไม่ขึ้นกับ $E\left(\hat{p}\right)$ ดังจะกล่าวถึงในหัวข้อถัดไป

**2.3 การแปลงแบบอาร์คไซน์**

กำหนดให้ $Y$ มีการแจกแจงแบบทวินาม มีพารามิเตอร์คือ $n$ และ $p$ และ $\hat{p} = Y/n$ คือค่าสัดส่วนตัวอย่าง ถ้าต้องการหาฟังก์ชัน $u(Y/n)$ ของ $Y/n$ ที่ทำให้มีค่าความแปรปรวนคงที่เมื่อ $n$ มีค่ามาก Hogg and Tanis (1993) ได้กล่าวไว้ว่าสามารถทำได้โดยใช้การกระจาย 2 เทอมของเทย์เลอร์ (Taylor's expansion) รอบค่า $p = E\left[\frac{Y}{n}\right]$ ดังนี้

$$u\left(\frac{Y}{n}\right) \approx u\left(E\left[\frac{Y}{n}\right]\right) + \left(\frac{Y}{n} - E\left[\frac{Y}{n}\right]\right)u'\left(E\left[\frac{Y}{n}\right]\right).$$

เมื่อพิจารณาค่าเฉลี่ยของ $u(Y/n)$ จะได้ว่า

$$E\left\{u\left(\frac{Y}{n}\right)\right\} \approx E\left\{u\left(E\left[\frac{Y}{n}\right]\right) + \left(\frac{Y}{n} - E\left[\frac{Y}{n}\right]\right)u'\left(E\left[\frac{Y}{n}\right]\right)\right\}$$

$$\approx u\left(E\left[\frac{Y}{n}\right]\right) = u(p)$$

ดังนั้นจะได้ว่า

$$u\left(\frac{Y}{n}\right) - E\left\{u\left(\frac{Y}{n}\right)\right\} \approx \left(\frac{Y}{n} - E\left[\frac{Y}{n}\right]\right)u'\left(E\left[\frac{Y}{n}\right]\right)$$

และพิจารณาค่าความแปรปรวนของ $u(Y/n)$ จะได้ว่า

$$Var\left\{u\left(\frac{Y}{n}\right)\right\} = E\left\{u\left(\frac{Y}{n}\right) - E\left[u\left(\frac{Y}{n}\right)\right]\right\}^2 \approx \left\{u'\left(E\left[\frac{Y}{n}\right]\right)\right\}^2 E\left(\frac{Y}{n} - E\left[\frac{Y}{n}\right]\right)^2$$

$$\approx \left\{u'\left(E\left[\frac{Y}{n}\right]\right)\right\}^2 Var\left(\frac{Y}{n}\right) \approx \{u'(p)\}^2 \frac{p(1-p)}{n}$$

และถ้า $Var\left[u\left(\frac{Y}{n}\right)\right]$ เป็นค่าคงที่ที่เกี่ยวเนื่องกับค่า $p$ แสดงให้เห็นว่า

$$\{u'(p)\}^2 p(1-p) = k_1 \quad \text{หรือ} \quad u'(p) = \frac{c_1}{\sqrt{p(1-p)}}$$

เมื่อ $k$ และ $c$ เป็นค่าคงที่ และจากค่าของ $u'(p)$ เมื่อต้องการหา $u(p)$ สามารถทำได้ดังนี้

$$u(p) = \int \frac{c_1}{\sqrt{p(1-p)}}\,dp = c_1 \int \frac{dp}{\sqrt{p(1-p)}}$$

ถ้ากำหนดให้ $t = \sqrt{p}$ จะได้ว่า $t^2 = p,\ 2t\,dt = dp$ และ

$$u(p) = c_1 \int \frac{dp}{\sqrt{p(1-p)}} = c_1 \int \frac{2t}{\sqrt{t^2(1-t^2)}}\,dt = 2c_1 \int \frac{1}{\sqrt{1-t^2}}\,dt$$

$$= 2c_1 \arcsin t + c_2 = 2c_1 \arcsin\sqrt{p} + c_2$$

เมื่อกำหนดให้ $c_1 = 1/2$ และ $c_2 = 0$ จะได้ว่า $u(p) = \arcsin\sqrt{p}$ ที่มีค่าความแปรปรวน

$$Var\{u(p)\} \approx Var\{\arcsin\sqrt{p}\}$$

$$\approx \left\{\frac{d}{dp}\arcsin\sqrt{p}\right\}^2 \frac{p(1-p)}{n}$$

$$\approx \left\{\frac{1}{2}\frac{1}{\sqrt{p(1-p)}}\right\}^2 \frac{p(1-p)}{n}$$

$$\approx \frac{1}{4}\frac{1}{p(1-p)}\frac{p(1-p)}{n} = \frac{1}{4n}$$

จากทฤษฎีบทลิมิตเข้าสู่ส่วนกลาง จะได้ว่า $\arcsin\sqrt{\hat{p}}$ สามารถประมาณด้วยการแจกแจงปกติที่มีค่าเฉลี่ย $\arcsin\sqrt{p}$ และค่าความแปรปรวน $1/4n$ แสดงว่าถ้า $n$ มีขนาดใหญ่

$$\arcsin\sqrt{\hat{p}} \sim \mathrm{N}\left(\arcsin\sqrt{p}, \frac{1}{4\mathrm{n}}\right) \qquad \text{(โดยประมาณ)}$$

และ

$$Z = \frac{\arcsin\sqrt{\hat{p}} - \arcsin\sqrt{p}}{1/2\sqrt{n}} \sim \mathrm{N}(0,1) \qquad \text{(โดยประมาณ)}$$

## 2.4 ความสัมพันธ์ระหว่างการแจกแจงทวินามกับการแจกแจงปัวส์ซอง

สมมติว่าเหตุการณ์ใดเหตุการณ์หนึ่งเกิดขึ้นในช่วงเวลา (0,t) ด้วยอัตราเฉลี่ย $\lambda$ ต่อหน่วยเวลา และเราสามารถแบ่งช่วงเวลา (0,t) นี้เป็นช่วงย่อยเล็ก ๆ $n$ ช่วงที่เป็นอิสระซึ่งกันและกัน โดยมีค่าความน่าจะเป็นของเหตุการณ์มากกว่าหนึ่งเหตุการณ์ที่เกิดขึ้นในแต่ละช่วงเวลาย่อยที่กำหนดมีค่าลู่ไปสู่ศูนย์

ให้ $Y$ คือจำนวนของลักษณะที่สนใจ ที่จะเกิดขึ้นในช่วงเวลา (0,t)

ถ้าต้องการหาความน่าจะเป็นที่จะเกิดลักษณะที่สนใจขึ้น $y$ ครั้งในช่วงเวลานี้ จะพบว่าคือความน่าจะเป็นที่จะเกิดลักษณะที่สนใจขึ้นในช่วงเวลาย่อย $y$ ช่วง ๆ ละ 1 ครั้ง จากทั้งหมด $n$ ช่วง ซึ่งในแต่ละช่วงเวลาย่อย ๆ หนึ่งนี้อาจมีลักษณะที่สนใจเกิดขึ้นหรือไม่ก็ได้ (เกิด-ไม่เกิด) แสดงว่าการเกิดลักษณะที่สนใจในแต่ละช่วงเวลาย่อยนี้มีการแจกแจงแบร์นูลลีที่มีพารามิเตอร์คือ $p = \lambda/n$ และเนื่องจากการเกิดลักษณะที่สนใจในแต่ละช่วงเวลาย่อย ๆ นี้เป็นอิสระต่อกัน ดังนั้นจำนวนรวมของลักษณะที่สนใจที่เกิดขึ้น $y$ ครั้งในช่วงเวลาย่อย $n$ ช่วง จึงมีการแจกแจงทวินามที่มีพารามิเตอร์ $n$ และ $p = \lambda/n$ นั่นก็คือ

จาก $P(Y=y) = \binom{\mathrm{n}}{\mathrm{y}} p^{y}(1-p)^{n-y}$ เนื่องจาก $p = \dfrac{\lambda}{n}$ จะพบว่า

$$= \frac{n!}{y!(n-y)!}\left(\frac{\lambda}{n}\right)^{y}\left(1-\frac{\lambda}{n}\right)^{n-y}$$

และถ้าแบ่งช่วงเวลาให้มากขึ้น $(n \to \infty)$ จะได้ว่า

$$\lim_{n\to\infty} \frac{n!}{y!(n-y)!}\left(\frac{\lambda}{n}\right)^{y}\left(1-\frac{\lambda}{n}\right)^{n-y} = \lim_{n\to\infty} \frac{n(n-1)..(n-y+1)}{n^{y}}\left(\frac{\lambda^{y}}{y!}\right)\left(1-\frac{\lambda}{n}\right)^{-y}\left(1-\frac{\lambda}{n}\right)^{n}$$

และเมื่อกำหนดค่า $y$ ให้คงที่ จะได้ว่า

$$\lim_{n\to\infty} \frac{n(n-1)..(n-y+1)}{n^{y}} = \lim_{n\to\infty}\left[(1)\left(1-\frac{1}{n}\right)..\left(1-\frac{y-1}{n}\right)\right] = 1$$

$$\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{n} = e^{-\lambda}$$

$$\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{-y} = 1$$

ดังนั้น $$\lim_{n\to\infty} \frac{n!}{y!(n-y)!}\left(\frac{\lambda}{n}\right)^{y}\left(1-\frac{\lambda}{n}\right)^{n-y} = \frac{\lambda^{y}e^{-\lambda}}{y!} = P(Y=y)$$

ซึ่งก็คือฟังก์ชันความน่าจะเป็นของการแจกแจงปัวส์ซองนั่นเอง

## 2.5 ความสัมพันธ์ระหว่างการแจกแจงปัวส์ซองกับการแจกแจงแกมมา

ในกระบวนการปัวส์ซองที่มีค่าเฉลี่ยเท่ากับ $\lambda$ เราทราบว่าระยะเวลาที่รอคอยจนกระทั่งเกิดความสำเร็จขึ้นเป็นครั้งแรก มีการแจกแจงแบบเอ็กโพเนนเชียล (Exponential distribution)

ถ้ากำหนดให้ $Y$ คือช่วงเวลาที่รอคอยจนกระทั่งเกิดความสำเร็จขึ้น $\alpha$ ครั้ง และเมื่อต้องการหาการแจกแจงของ $Y$ สามารถทำได้ดังนี้

ถ้า $y > 0$ ฟังก์ชันการแจกแจงของสะสมของ $Y$ กำหนดโดย

$$F(y) = P(Y \le y) = 1 - P(Y > y)$$

= 1 - P( เกิดความสำเร็จน้อยกว่า $\alpha$ ครั้ง ในช่วง $[0, y]$ )

$$= 1 - \left\{ e^{-\lambda y} + \lambda y e^{-\lambda y} + \ldots + \frac{1}{(\alpha-2)!}(\lambda y)^{\alpha-2} e^{-\lambda y} + \frac{1}{(\alpha-1)!}(\lambda y)^{\alpha-1} e^{-\lambda y} \right\}$$

$$= 1 - \sum_{k=0}^{\alpha-1} \frac{(\lambda y)^k e^{-\lambda y}}{k!}$$

เนื่องจากจำนวนครั้งของความสำเร็จช่วง $[0, y]$ มีการแจกแจงปัวส์ซองที่มีค่าเฉลี่ย $\lambda y$ และเนื่องจากลักษณะของ $Y$ เป็นตัวแปรสุ่มชนิดต่อเนื่อง จึงสามารถหาฟังก์ชันความหนาแน่นได้จากการหาอนุพันธ์อันดับที่ 1 ซึ่งจะได้ว่า

$$F'(y) = \mathrm{f(y)} = \lambda \mathrm{e}^{-\lambda y} - e^{-\lambda y} \sum_{k=1}^{\alpha-1} \left[ \frac{k(\lambda y)^{k-1} \lambda}{k!} - \frac{(\lambda y)^{k-1} \lambda}{k!} \right]$$

$$= \lambda \mathrm{e}^{-\lambda y} - e^{-\lambda y} \left[ \lambda - \frac{\lambda(\lambda y)^{\alpha-1}}{(\alpha-1)!} \right]$$

$$= \frac{\lambda(\lambda y)^{\alpha-1}}{(\alpha-1)!} e^{-\lambda y}$$

พิจารณาฟังก์ชันที่ได้ ก็คือฟังก์ชันความหนาแน่นของการแจกแจงแกมมาซึ่งมีชื่อเรียกเฉพาะว่า การแจกแจงเออร์แรง (Erlang distribution) ในกรณีที่ $\alpha$ เป็นจำนวนเต็มบวก

### 2.6 ความสัมพันธ์ระหว่างสถิติอันดับกับการแจกแจงบีตา

กำหนดให้ $X_{(1)} < X_{(2)} < \ldots < X_{(n)}$ คือสถิติอันดับของตัวแปรเชิงสุ่ม $X_1, X_2 \ldots, X_n$ ซึ่งมาจากประชากรที่มีการแจกแจงแบบต่อเนื่อง โดยมีฟังก์ชันสะสม $F_X(x)$ และฟังก์ชันความหนาแน่น $f_X(x)$ แล้ว จะได้ว่าฟังก์ชันความหนาแน่นของ $X_{(j)}$ คือ

$$f_{X_{(j)}}(x)=\frac{n!}{(j-1)!(n-j)!}f_X(x)[F_X(x)]^{j-1}[1-F_X(x)]^{n-j}$$

ดังนั้นถ้าตัวแปรสุ่ม $X_1, X_2, \ldots, X_n$ มีการแจกแจงแบบเอกรูปในช่วง (0,1) จะได้ว่า $f_X(x)=1$ เมื่อ $x \in (0,1)$ และ $F_X(x)=x$ เมื่อ $x \in (0,1)$ และเมื่อต้องการหาฟังก์ชันความหนาแน่นของสถิติอันดับที่ $j$ จะได้ว่า

$$\begin{aligned} f_{X_{(j)}}(x) &= \frac{n!}{(j-1)!(n-j)!}x^{j-1}(1-x)^{n-j} \\ &= \frac{\Gamma(n+1)}{\Gamma(j)\Gamma(n-j+1)}x^{j-1}(1-x)^{(n-j+1)-1} \end{aligned}$$

ซึ่งก็คือฟังก์ชันความหนาแน่นของการแจกแจงบีตาที่มีพารามิเตอร์คือ $j$ และ $n-j+1$ นั่นก็คือตัวสถิติอันดับที่ $j$ จากตัวแปรสุ่มที่มีการแจกแจงเอกรูปในช่วง (0,1) จะมีการแจกแจงแบบบีตาที่มีพารามิเตอร์คือ $j$ และ $n-j+1$

## 2.7 ความสัมพันธ์ระหว่างการแจกแจงบีตากับการแจกแจงเอฟ

Hald (1952) ได้อธิบายความสัมพันธ์ระหว่างฟังก์ชันบีตาที่ไม่สมบูรณ์กับการแจกแจงเอฟดังนี้

ถ้า B คือตัวแปรสุ่มบีตาที่มีพารามิเตอร์ $(\upsilon_1, \upsilon_2)$ เราสามารถสร้างตัวแปรสุ่มเอฟที่มีพารามิเตอร์ $2\upsilon_1$ และ $2\upsilon_2$ ได้จากความสัมพันธ์ดังต่อไปนี้

$$f=\frac{\upsilon_2 x}{\upsilon_1(1-x)} \quad ;(0 \le x \le 1)$$

โดย $x$ คือค่าที่สอดคล้องกับ $\mathrm{I}_X(\upsilon_1, \upsilon_2)$ ซึ่งก็คือฟังก์ชันความน่าจะเป็นสะสมของตัวแปรสุ่มบีตา หรือที่เรียกว่าอัตราส่วนของฟังก์ชันบีตาที่ไม่สมบูรณ์ (Incomplete Beta function ratio) ซึ่งมีรูปแบบดังนี้

$$F(x) = \frac{1}{B(\upsilon_1, \upsilon_2)} \int_0^X w^{\upsilon_1 - 1}(1-w)^{\upsilon_2 - 1} dw$$

$$= \frac{B_X(\upsilon_1, \upsilon_2)}{B(\upsilon_1, \upsilon_2)}$$

$$= I_X(\upsilon_1, \upsilon_2)$$

โดยที่

$$B(y, n-y+1) = \int_0^p w^{y-1}(1-w)^{n-y} dw$$

คือ ฟังก์ชันบีตา (Beta function)

$$B(y, n-y+1) = \int_0^1 w^{y-1}(1-w)^{n-y} dw$$

คือ ฟังก์ชันบีตาที่ไม่สมบูรณ์ (Incomplete Beta function)

## 2.8 การประมาณค่าพารามิเตอร์แบบช่วง

การประมาณค่าแบบช่วงหรือที่เรียกว่า ช่วงความเชื่อมั่น (confidence intervals) เป็นการประมาณค่าพารามิเตอร์ของประชากรว่าอยู่ในช่วงใดช่วงหนึ่งโดยใช้ข้อมูลตัวอย่าง และช่วงของการประมาณค่าจะบอกถึงค่าที่ต่ำสุดและค่าที่สูงสุดของพารามิเตอร์ นั่นคือ

ให้ $X_1, X_2, ..., X_n$ เป็นตัวแปรเชิงสุ่มจากการแจกแจงซึ่งมี $\theta$ เป็นพารามิเตอร์ ให้ $t_1(X_1, ..., X_n), t_2(X_1, ..., X_n)$ เป็นตัวสถิติที่ $t_1 < t_2$ และ

$$P\{ t_1 < \theta < t_2 \} = 1 - \alpha$$

ช่วงสุ่ม (random interval) ของ $(t_1, t_2)$ ที่ได้ก็คือ ตัวประมาณค่าแบบช่วง (interval estimator)

โดยที่ $t_1$ และ $t_2$ **คือขีดจำกัดความเชื่อมั่นล่าง** (lower confidence limit) **และขีดจำกัดความเชื่อมั่นบน** (upper confidence limit) · **สำหรับพารามิเตอร์** $\theta$ และ $1-\alpha$ เรียกว่า **สัมประสิทธิ์ความเชื่อมั่น** (confidence coefficient) **หรือระดับความเชื่อมั่น** (confidence level) โดยทั่วไปมักจะกำหนดค่าระดับความเชื่อมั่นเป็น 90%, 95% และ 99%

ยกตัวอย่างเช่น $P\{ t_1 < \theta < t_2 \} = 0.90$ หมายถึง ความน่าจะเป็นที่ค่าของ $\theta$ จะมีค่าอยู่ในช่วงของ $t_1$ และ $t_2$ มีค่าเท่ากับ 0.90 หรือ 90% และความน่าจะเป็นที่ค่าของ $\theta$ จะมีค่าน้อยกว่า $t_1$ หรือมีค่ามากกว่า $t_2$ มีค่าเท่ากับ 0.10 หรือ 10%

นอกจากต้องพิจารณาค่าระดับความเชื่อมั่นแล้ว สิ่งจำเป็นที่ควรพิจารณาถึงสำหรับการประมาณค่าแบบช่วงคือ ความยาวเฉลี่ยของช่วงประมาณ ซึ่งคือค่าคาดหวังของผลต่างของขีดจำกัดความเชื่อมั่นบนและขีดจำกัดความเชื่อมั่นล่าง หรือหาได้จากค่า $E(t_2 - t_1)$

## 2.9 <u>การประมาณค่าสัดส่วนประชากรแบบช่วง</u>

ในการหาค่าประมาณแบบช่วงสำหรับค่าสัดส่วนประชากร ที่ระดับความเชื่อมั่น $(1-\alpha)100\%$ เมื่อกำหนดค่า $p$ ใด ๆ โดยที่ $0 < p < 1$ Larson (1982) ได้อธิบายถึงวิธีการประมาณค่าสัดส่วนประชากรแบบช่วงไว้ดังนี้

ถ้ากำหนดให้ $h_1(p)$ และ $h_2(p)$ คือขีดจำกัดความเชื่อมั่นล่างและขีดจำกัดความเชื่อมั่นบนของ $p$ ที่ระดับความเชื่อมั่น $(1-\alpha)100\%$ แล้วจะได้ว่า

$$P\{h_1(p) < p < h_2(p) \mid \mathrm{Y} = \mathrm{y}\} = 1 - \alpha/2$$

นั่นก็คือในการประมาณค่า $p$ ที่ระดับความเชื่อมั่น $(1-\alpha)100\%$ สามารถหาได้จากค่าจำนวนเต็มที่มีค่ามากที่สุดของ $h_1(p)$ ที่ทำให้ $P\{Y \le h_1(p)\} \le \alpha/2$ และค่าจำนวนเต็มที่เล็กที่สุดของ $h_2(p)$ ที่ทำให้ $P\{Y \ge h_2(p)\} \le \alpha/2$ เนื่องจากการแจกแจงทวินามเป็นการแจกแจงที่ไม่ต่อเนื่องดังนั้นจึงไม่สามารถหาค่าให้เท่ากับ $\alpha/2$ ได้

และเนื่องจากเหตุการณ์ $\{Y \le h_1(p)\}$ และ $\{Y \ge h_2(p)\}$ เป็นเหตุการณ์ที่แยกจากกันโดยเด็ดขาด (mutually disjoint) จะได้ว่า

$$P\{Y \le h_1(p) \text{ หรือ } Y \ge h_2(p)\} \le \frac{\alpha}{2} + \frac{\alpha}{2} = \alpha$$

ดังนั้นเมื่อพิจารณาส่วนเติมเต็มของเหตุการณ์ จะพบว่ามีค่าความน่าจะเป็นอย่างน้อยที่สุด $1-\alpha$ นั่นก็คือ

$$P\{h_1(p) \le Y \le h_2(p)\} = 1 - P\{Y \le h_1(p) \text{ หรือ } Y \ge h_2(p)\}$$
$$\ge 1 - \alpha$$

เมื่อพิจารณาฟังก์ชัน $h_1(p)$ และ $h_2(p)$ เนื่องจากเป็นฟังก์ชันทางเดียวที่ไม่ลดลงจึงสามารถหาฟังก์ชันผกผัน (inverse function) ของเหตุการณ์ได้ นั่นคือเมื่อ $Y = h_1(p)$ จะได้ว่า $p = h_1^{-1}(Y)$ ฟังก์ชัน $h_2(p)$ ก็เช่นเดียวกันดังนั้นจึงได้ว่าเหตุการณ์ $P\{h_1(p) \le Y \le h_2(p)\}$ ก็คือเหตุการณ์ $\mathrm{P}\{h_2^{-1}(Y) \le p \le h_1^{-1}(Y)\}$ ดังนั้น

$$\mathrm{P}\{h_2^{-1}(Y) \le p \le h_1^{-1}(Y)\} \ge 1-\alpha$$

แสดงว่า $\{h_2^{-1}(Y)\ ,\ h_1^{-1}(Y)\}$ คือค่าประมาณแบบช่วงของ $p$ ที่ระดับความเชื่อมั่น $(1-\alpha)100\%$

เมื่อพิจารณา $h_1(p)$ เนื่องจากคือจำนวนเต็มที่มีค่ามากที่สุดที่ทำให้ $P\{Y \le h_1(p)\} \le \alpha/2$ ดังนั้นจึงได้ว่า

$$\sum_{j=o}^{h_1(p)} \binom{n}{j} p^j (1-p)^{n-j} \le \alpha/2$$

สมมติได้ค่าสังเกตคือ $Y = y$ จะได้ว่า $h_1^{-1}(y)$ คือขีดจำกัดความเชื่อมั่นบนของ $p$ และเนื่องจาก $h_1(h_1^{-1}(y)) = y$ ดังนั้นการหาขีดจำกัดความเชื่อมั่นบนของ $p$ คือการหาค่า $p$ ที่ทำให้

$$\sum_{j=o}^{y} \binom{n}{j} p^j (1-p)^{n-j} = \alpha/2$$

ทำนองเดียวกัน $h_2(h_2^{-1}(y)) = y$ ดังนั้นขีดจำกัดความเชื่อมั่นล่างของ $p$ ก็คือค่าของ $p$ ที่ทำให้

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

ดังนั้นสามารถสรุปได้ว่า ถ้า $Y$ เป็นตัวแปรสุ่มที่มีการแจกแจงแบบทวินาม ที่มีพารามิเตอร์ $n$ และ $p$ ให้ $p_L$ และ $p_U$ คือค่าที่ทำให้

$$\sum_{j=y}^{n}\binom{n}{j}p_L{}^{j}(1-p_L)^{n-j} = \alpha/2$$

$$\sum_{j=o}^{y}\binom{n}{j}p_U{}^{j}(1-p_U)^{n-j} = \alpha/2$$

สำหรับค่าสังเกต $y$ ใด ๆ แล้ว จะได้ว่า $(p_L, p_U)$ คือค่าประมาณสัดส่วนประชากรแบบช่วง ที่ระดับความเชื่อมั่น $(1-\alpha)100\%$

จากสูตรการหาขีดจำกัดความเชื่อมั่นสำหรับค่าสัดส่วนประชากร ซึ่งต้องอาศัยการแก้สมการและคำนวณค่าแฟคทอเรียล พบว่าถ้า $n$ มีค่ามากจะประสบปัญหาในการประมาณค่า ดังนั้นจึงต้องอาศัยการประมาณโดยวิธีการอื่น ๆ ดังเช่น วิธีปกติ วิธีแปลงแบบอาร์คไซน์ วิธีสคอร์ วิธีปัวส์ซอง หรือวิธีเอฟ เป็นต้น ซึ่งแต่ละวิธีประมาณมีรายละเอียดต่อไปนี้

### 2.9.1 วิธีปกติ

จากทฤษฎีบทลิมิตสู่ส่วนกลางกล่าวว่า ถ้าตัวอย่างมีขนาดใหญ่ ค่าสัดส่วนตัวอย่าง $\hat{p}$ จะมีการแจกแจงโดยประมาณแบบปกติที่มีค่าเฉลี่ย $p$ และความแปรปรวน $p(1-p)/n$ นั่นคือเมื่อ $n \to \infty$

$$\hat{p} \sim N\left( p , \frac{p(1-p)}{n} \right) \qquad \text{(โดยประมาณ)}$$

และได้ว่า

$$Z = \frac{\hat{p}-p}{\sqrt{p(1-p)/n}} \sim N(0,1) \qquad \text{(โดยประมาณ)}$$

ดังนั้นในการประมาณค่าสัดส่วนประชากรแบบช่วง จึงสามารถวิเคราะห์ได้จากความสัมพันธ์

$$P\left(-Z_{1-\alpha/2} < Z < Z_{1-\alpha/2}\right) = 1-\alpha$$

$$P\left(-Z_{1-\alpha/2} < \frac{\hat{p}-p}{\sqrt{p(1-p)/n}} < Z_{1-\alpha/2}\right) = 1-\alpha$$

$$P\left(\hat{p}-Z_{1-\alpha/2}\sqrt{\frac{p(1-p)}{n}} < p < \hat{p}+Z_{1-\alpha/2}\sqrt{\frac{p(1-p)}{n}}\right) = 1-\alpha$$

เนื่องจากไม่ทราบค่า $p$ จึงไม่ทราบค่าความแปรปรวน $p(1-p)/n$ ดังนั้นจะประมาณด้วยค่าความแปรปรวนตัวอย่าง คือ $\hat{p}\left(1-\hat{p}\right)/n$ และได้ค่าประมาณแบบช่วงของ $p$ ที่ระดับความเชื่อมั่น $(1-\alpha)100\%$ คือ

$$\left(\hat{p}-Z_{1-\alpha/2}\sqrt{\hat{p}\left(1-\hat{p}\right)/n}\ ,\ \hat{p}+Z_{1-\alpha/2}\sqrt{\hat{p}\left(1-\hat{p}\right)/n}\right)$$

**2.9.2 วิธีแปลงแบบอาร์คไซน์**

เนื่องจากถ้าตัวอย่างมีขนาดใหญ่ค่าสัดส่วนตัวอย่าง $\hat{p}$ จะมีการแจกแจงโดยประมาณแบบปกติที่มีค่าเฉลี่ย $p$ และความแปรปรวน $p(1-p)/n$ และโดยใช้หลักการแปลงเชิงเส้น Johnson and Kotz (1969) ได้แนะนำการแปลงแบบอาร์คไซน์ ซึ่ง Ghosh (1979) ได้นำมาใช้ในการประมาณค่าสัดส่วนประชากรแบบช่วง ซึ่งได้ว่า $\arcsin\sqrt{\hat{p}}$ จะมีการแจกแจงโดยประมาณแบบปกติที่มีค่าเฉลี่ย $\arcsin\sqrt{p}$ และความแปรปรวน $1/4n$ ดังนั้นจะได้ว่า

$$Z = \frac{\arcsin\sqrt{\hat{p}} - \arcsin\sqrt{p}}{1/2\sqrt{n}} \sim N(0,1) \qquad \text{(โดยประมาณ)}$$

และการประมาณค่าสัดส่วนประชากรแบบช่วงสามารถหาได้จากความสัมพันธ์ดังนี้

$$P\left(-Z_{1-\alpha/2} < \frac{\arcsin\sqrt{\hat{p}} - \arcsin\sqrt{p}}{1/2\sqrt{n}} < Z_{1-\alpha/2}\right) = 1-\alpha$$

โดยจะได้ค่าประมาณแบบช่วงของ $\arcsin\sqrt{p}$ คือ

$$\arcsin\sqrt{\hat{p}} - \frac{Z_{1-\alpha/2}}{2\sqrt{n}} < \arcsin\sqrt{p} < \arcsin\sqrt{\hat{p}} + \frac{Z_{1-\alpha/2}}{2\sqrt{n}}$$

พิจารณาขีดจำกัดความเชื่อมั่นล่าง เนื่องจาก

$$\arcsin\sqrt{\hat{p}} - \frac{Z_{1-\alpha/2}}{2\sqrt{n}} < \arcsin\sqrt{p}$$

แสดงว่า

$$\sin\left(\arcsin\sqrt{\hat{p}} - \frac{Z_{1-\alpha/2}}{2\sqrt{n}}\right) < \sqrt{p}$$

นั่นก็คือ

$$\sin^2\left(\arcsin\sqrt{\hat{p}} - \frac{Z_{1-\alpha/2}}{2\sqrt{n}}\right) < p$$

เช่นเดียวกับการหาขีดจำกัดความเชื่อมั่นบน ดังนั้นจะได้ค่าประมาณแบบช่วงของ $p$ คือ

$$\left(\sin^2\left[\arcsin\sqrt{\hat{p}} - \frac{Z_{1-\alpha/2}}{2\sqrt{n}}\right], \sin^2\left[\arcsin\sqrt{\hat{p}} + \frac{Z_{1-\alpha/2}}{2\sqrt{n}}\right]\right)$$

### 2.9.3 วิธีสคอร์

Ghosh (1979) และ Chen (1990) ได้เคยศึกษาวิธีประมาณค่าสัดส่วนประชากรแบบช่วงโดยใช้วิธีการของเบย์ (Bayes method) และพบว่าวิธีประมาณดังกล่าวสามารถสร้างได้จากสถิติที่ใช้สำหรับการทดสอบสมมติฐาน ซึ่งต่อมา Alan and Brent (1998) ได้ทำการศึกษาวิธีประมาณนี้และเรียกว่าวิธีสคอร์ โดยมีหลักในการสร้างตัวประมาณดังนี้

สำหรับการทดสอบสมมติฐาน $H_o : p = p_0$ เทียบกับ $H_1 : p \neq p_0$ ที่ระดับนัยสำคัญ $\alpha$ พบว่าเราจะยอมรับสมมติฐานหลักเมื่อ

$$\frac{\left|\hat{p} - p_0\right|}{\sqrt{p_0(1-p_0)/n}} \leq Z_{1-\alpha/2}$$

และเมื่อยกกำลังสองจะได้สถิติทดสอบที่มีการแจกแจงไค-สแคว์ มีองศาความเป็นอิสระเท่ากับ 1 ซึ่งมีรูปแบบดังนี้

$$\left\{\frac{\hat{p} - \mathrm{p}_0}{\sqrt{\mathrm{p}_0(1-p_0)/n}}\right\}^2 = \chi_1^2$$

เมื่อกระจายเทอมกำลังสองสามารถเขียนได้ในรูป

$$\left(\hat{p} - p_0\right)^2 \leq Z_{1-\alpha/2}^2\left\{\sqrt{\frac{p_0(1-p_0)}{n}}\right\}^2$$

$$\hat{p}^2 - 2\hat{\mathrm{p}}\mathrm{p}_0 + p_0^2 - Z_{1-\alpha/2}^2\frac{\mathrm{p}_0}{n} + Z_{1-\alpha/2}^2\frac{\mathrm{p}_0^2}{n} \leq 0$$

$$\hat{p}^2 - p_0\left(2\hat{p} + \frac{Z_{1-\alpha/2}^2}{n}\right) + p_0^2\left(1 + \frac{Z_{1-\alpha/2}^2}{n}\right) \leq 0$$

ในทำนองเดียวกันสามารถเขียนได้ในรูป (โดยให้ $p$ แทน $p_0$)

$$\hat{p}^2 - p\left(2\hat{p} + \frac{Z_{1-\alpha/2}^2}{n}\right) + p^2\left(1 + \frac{Z_{1-\alpha/2}^2}{n}\right) \leq 0$$

และเมื่อกำหนดให้สมการมีค่าเท่ากับ 0 จะได้ว่า

$$p^2\left(1+\frac{Z^2_{1-\alpha/2}}{n}\right)-p\left(2\hat{p}+\frac{Z^2_{1-\alpha/2}}{n}\right)+\hat{p}^2 \;=\; 0$$

**ก็จะอยู่ในรูปของสมการกำลังสอง (quadratic equation) คือ** $ax^2+bx+c = 0$ **เมื่อ**

$$a=\left(1+\frac{Z^2_{1-\alpha/2}}{n}\right),\;\; \mathrm{b}=-\left(2\hat{p}+\frac{Z^2_{1-\alpha/2}}{n}\right),\;\; \mathrm{c}=\hat{p}^2 \;\;\text{และ}\;\; x=p$$

และสามารถหารากของสมการกำลังสองได้จากสูตร $x=\dfrac{-b\pm\sqrt{b^2-4ac}}{2a}$ นั่นคือ

$$p \;=\; \frac{\left(2\hat{p}+\frac{Z^2_{1-\alpha/2}}{n}\right)\pm\sqrt{\left(2\hat{p}+\frac{Z^2_{1-\alpha/2}}{n}\right)^2-4\left(1+\frac{Z^2_{1-\alpha/2}}{n}\right)\hat{p}^2}}{2\left(1+\frac{Z^2_{1-\alpha/2}}{n}\right)}$$

เนื่องจาก

$$\sqrt{\left(2\hat{p}+\frac{Z^2_{1-\alpha/2}}{n}\right)^2-4\left(1+\frac{Z^2_{1-\alpha/2}}{n}\right)\hat{p}^2} \;=\; \sqrt{\frac{Z^2_{1-\alpha/2}}{n}\left\{4\hat{p}\left(1-\hat{p}\right)+\frac{Z^2_{1-\alpha/2}}{n}\right\}}$$

$$=\; \frac{Z_{1-\alpha/2}}{\sqrt{n}}\sqrt{4\hat{p}\left(1-\hat{p}\right)+\frac{Z^2_{1-\alpha/2}}{n}}$$

แสดงว่า

$$p \;=\; \frac{2\hat{p}+\frac{Z^2_{1-\alpha/2}}{n}\pm\frac{Z_{1-\alpha/2}}{\sqrt{n}}\sqrt{4\hat{p}\left(1-\hat{p}\right)+\frac{Z^2_{1-\alpha/2}}{n}}}{2\left(1+\frac{Z^2_{1-\alpha/2}}{n}\right)}$$

หรือ

$$p = \frac{\hat{p}+\dfrac{Z_{1-\alpha/2}^{2}}{2n} \pm \dfrac{Z_{1-\alpha/2}}{2\sqrt{n}}\sqrt{4\hat{p}\left(1-\hat{p}\right)+\dfrac{Z_{1-\alpha/2}^{2}}{n}}}{1+\dfrac{Z_{1-\alpha/2}^{2}}{n}}$$

$$= \frac{\hat{p}+\dfrac{Z_{1-\alpha/2}^{2}}{2n} \pm Z_{1-\alpha/2}\sqrt{\dfrac{\hat{p}\left(1-\hat{p}\right)}{n}+\dfrac{Z_{1-\alpha/2}^{2}}{4n^{2}}}}{1+\dfrac{Z_{1-\alpha/2}^{2}}{n}}$$

$$= \frac{\hat{p}+\dfrac{Z_{1-\alpha/2}^{2}}{2n} \pm Z_{1-\alpha/2}\sqrt{\left\{\hat{p}\left(1-\hat{p}\right)+\dfrac{Z_{1-\alpha/2}^{2}}{4n}\right\}\Big/n}}{1+\dfrac{Z_{1-\alpha/2}^{2}}{n}}$$

จะได้สถิติที่สามารถนำมาใช้ในการหาค่าประมาณสัดส่วนประชากรแบบช่วง ซึ่งมีรูปแบบดังนี้

$$\left(\frac{\hat{p}+\dfrac{Z_{1-\alpha/2}^{2}}{2n}-Z_{1-\alpha/2}\sqrt{\left[\hat{p}(1-\hat{p})+Z_{1-\alpha/2}^{2}/4n\right]\Big/n}}{1+\dfrac{Z_{1-\alpha/2}^{2}}{n}},\frac{\hat{p}+\dfrac{Z_{1-\alpha/2}^{2}}{2n}+Z_{1-\alpha/2}\sqrt{\left[\hat{p}(1-\hat{p})+Z_{1-\alpha/2}^{2}/4n\right]\Big/n}}{1+\dfrac{Z_{1-\alpha/2}^{2}}{n}}\right)$$



**2.9.4 วิธีปัวส์ซอง**

เนื่องจากเราสามารถประมาณขีดจำกัดความเชื่อมั่นล่าง ของค่าสัดส่วนประชากรสำหรับตัวแปรสุ่มทวินาม ได้จากสมการ

$$\sum_{j=y}^{n}\binom{n}{j}p_{L}^{\,j}(1-p_{L})^{n-j} = \alpha/2$$

และเมื่อประมาณด้วยการแจกแจงปัวส์ซองจึงสามารถหาขีดจำกัดความเชื่อมั่นล่างโดยพิจารณาได้จากสมการ

$$\sum_{j=y}^{\infty} \frac{(\mathrm{np}_L)^j e^{-np_L}}{\mathrm{j}!} = \alpha/2$$

หรือ

$$1-\sum_{j=0}^{y-1} \frac{(\mathrm{np}_L)^j e^{-np_L}}{\mathrm{j}!} = \alpha/2$$

เมื่อพิจารณา $1-\sum_{j=0}^{y-1} \frac{(\mathrm{np}_L)^j e^{-np_L}}{\mathrm{j}!}$ จะพบว่าคือสูตรการหาฟังก์ชันการแจกแจงสะสมของการแจกแจงเออร์แรง ที่มีพารามิเตอร์คือ $np_L$ และ $y$ นั่นเอง

และเนื่องจาก $2np_L y$ คือตัวแปรสุ่ม $\chi^2$ ที่มีองศาความเป็นอิสระ $2y$ ดังนั้นในการหาค่าประมาณแบบช่วง จึงสามารถพิจารณาได้จาก

$$P\left(\chi^2_{2y} \le 2\mathrm{np}_L\right) = \alpha/2$$

และเนื่องจาก

$$P\left(\chi^2_{2y} \le \chi^2_{2y,\alpha/2}\right) = \alpha/2$$

ซึ่งจะได้ขีดจำกัดความเชื่อมั่นล่างคือ

$$p_L = \frac{1}{2\mathrm{n}} \chi^2_{2y,\alpha/2}$$

ทำนองเดียวกันก็จะได้ขีดจำกัดความเชื่อมั่นบนคือ

$$p_U = \frac{1}{2\mathrm{n}} \chi^2_{2(y+1),1-\alpha/2}$$

และจะได้ค่าประมาณแบบช่วงของค่าสัดส่วนประชากร คือ

$$\left(\frac{1}{2\mathrm{n}} \chi^2_{2y,\alpha/2} \ , \ \frac{1}{2\mathrm{n}} \chi^2_{2(y+1),1-\alpha/2}\right)$$

### 2.9.5 วิธีเอฟ

การประมาณค่าสัดส่วนประชากรโดยใช้การแจกแจงเอฟ Blyth (1986) ได้แสดงวิธีประมาณค่าโดยอาศัยความสัมพันธ์ระหว่างสถิติอันดับกับการแจกแจงแบบบีตา ดังนี้

ถ้า $X_{(1)} < X_{(2)} < \ldots < X_{(n)}$ คือสถิติอันดับของตัวแปรสุ่ม $X_1, X_2, \ldots, X_n$ ที่มีการแจกแจงแบบเอกรูปในช่วง (0,1) โดยอาศัยคุณสมบัติของสถิติอันดับ (หัวข้อ 2.6) จะได้ว่า $X_{(y)}$ มีการแจกแจงแบบบีตาที่มีพารามิเตอร์ $y$ และ $n-y+1$ ถ้ากำหนดให้ $Y$ คือจำนวนของ $X_i$ ซึ่งมีค่าน้อยกว่า $p$ ดังนั้น $Y$ จึงมีการแจกแจงทวินามที่มีพารามิเตอร์ $n$ และ $p$ เนื่องจากเหตุการณ์ $Y \geq y$ ก็คือเหตุการณ์ $X_{(y)} \leq p$ ดังนั้นจึงสามารถหาค่า $P(Y \geq y)$ โดยพิจารณาจากการหาค่า $P(X_{(y)} \langle p)$ ได้ซึ่งสามารถนำมาใช้ประโยชน์ในการหาขีดจำกัดความเชื่อมั่นล่างของค่าประมาณแบบช่วงได้ดังนี้

เนื่องจาก $P(Y \geq y) \;=\; P(X_{(y)} \langle p)$

$$= \frac{\Gamma(n+1)}{\Gamma(y)\Gamma(n-y+1)} \int_0^p x^{y-1}(1-x)^{n-y}\,dx$$

กำหนดให้ $t = (n-y+1)x/y(1-x)$ เมื่อนำไปแทนลงในสมการข้างบน จะได้

$$P(Y \geq y) = \frac{\Gamma(n+1)}{\Gamma(y)\Gamma(n-y+1)}\left(\frac{n-y+1}{y}\right)^{n-y+1}$$

$$\times \int_0^{(n-y+1)p/y(1-p)} \frac{t^{y-1}}{(\frac{n-y+1}{y}+t)}\,dt$$

$$= P\left(F_{2y,2(n-y+1)} \langle \frac{(n-y+1)p}{y(1-p)}\right)$$

โดยอาศัยความสัมพันธ์ระหว่างการแจกแจงเอฟกับการแจกแจงบีตา Johnson and Kotz (1970) กล่าวไว้ว่าสามารถหาค่าขีดจำกัดความเชื่อมั่นล่างที่ $(\alpha/2)100\%$ ของตัวแปรสุ่มเอฟที่มีองศาความเป็นอิสระ $2y$ และ $2(n-y+1)$ ได้ดังสมการ

$$I_{2yF/2(n-y+1)+2yF}\{y, n-y+1\} = \alpha/2$$

ดังนั้น การหาขีดจำกัดความเชื่อมั่นล่างสำหรับค่าสัดส่วนประชากร สามารถเขียนได้ในรูปของฟังก์ชันบีตา ดังนี้

$$I_{p_L}\{y, n-y+1\} = \alpha/2$$

จะได้ว่า

$$p_L = \frac{2yF_{2y,2(n-y+1),\alpha/2}}{2(n-y+1)+2yF_{2y,2(n-y+1),\alpha/2}}$$

ซึ่งสามารถเขียนได้ในรูปของ

$$p_L = \frac{y}{y+(n-y+1)F_{2(n-y+1),2y,1-\alpha/2}}$$

ทำนองเดียวกันก็จะได้ขีดจำกัดความเชื่อมั่นบนคือ

$$p_U = \frac{(y+1)F_{2(y+1),2(n-y),1-\alpha/2}}{n-y+(y+1)F_{2(y+1),2(n-y),1-\alpha/2}}$$

และได้ค่าประมาณแบบช่วงที่พิจารณาจากการแจกแจงเอฟคือ

$$\left(\frac{y}{y+(n-y+1)F_{2(n-y+1),2y,1-\alpha/2}} \; , \; \frac{(y+1)F_{2(y+1),2(n-y),1-\alpha/2}}{n-y+(y+1)F_{2(y+1),2(n-y),1-\alpha/2}}\right)$$

## 2.9 เกณฑ์การเปรียบเทียบ

สำหรับการเปรียบเทียบค่าประมาณสัดส่วนประชากรแบบช่วงจากวิธีการประมาณทั้ง 5 วิธีนั้น จะทำการเปรียบเทียบค่าระดับความเชื่อมั่นและความยาวเฉลี่ยของค่าประมาณแบบช่วง ที่คำนวณได้จากแต่ละสถานการณ์ทดลอง ในการทดลองซ้ำ 2,000 ครั้ง โดยมีรายละเอียดการเปรียบเทียบดังนี้

การเปรียบเทียบค่าระดับความเชื่อมั่น เพื่อตรวจสอบว่าวิธีการประมาณใดให้ค่าระดับความเชื่อมั่นไม่ต่ำกว่าระดับความเชื่อมั่นที่กำหนดนั้น เกณฑ์ที่ใช้ในการเปรียบเทียบจะอาศัยการทดสอบสมมติฐานโดยใช้ตัวสถิติ Z มีรูปแบบดังนี้

$$H_0 : c \geq c_0$$
$$H_1 : c < c_0$$

ซึ่งจะได้ว่า

$$-Z_{1-\alpha_0} < \frac{\hat{c} - c_0}{\sqrt{\frac{c_0(1-c_0)}{n}}} < 1$$

$$c_0 - Z_{1-\alpha_0}\sqrt{\frac{c_0(1-c_0)}{n}} < \hat{c} < 1$$

ฉะนั้น จะได้ช่วงของการยอมรับสมมติฐานหลักคือ

$$\left(c_0 - Z_{1-\alpha_0}\sqrt{\frac{c_0(1-c_0)}{n}}\ ,\ 1\right)$$

ดังนั้นจะได้ว่า วิธีการประมาณให้ค่าระดับความเชื่อมั่นไม่ต่ำกว่าเกณฑ์ที่กำหนดถ้า $\hat{c}$ มีค่าอยู่ในช่วง

$$\left(c_0 - Z_{1-\alpha_0}\sqrt{\frac{c_0(1-c_0)}{n}}\ ,\ 1\right)$$

โดยที่ $\alpha_0$ คือระดับนัยสำคัญหรือความน่าจะเป็นที่จะเกิดความผิดพลาดที่กำหนดในการทดสอบ สำหรับการวิจัยครั้งนี้กำหนดระดับนัยสำคัญของการทดสอบเท่ากับ 0.05

$c$ คือระดับความเชื่อมั่น

$\hat{c}$ คือระดับความเชื่อมั่นที่ได้จากการทดลอง

$c_0$ คือระดับความเชื่อมั่นที่กำหนด ( 0.90, 0.95 และ 0.99 )

$n$ จำนวนครั้งของการทดลอง (2,000)

**1. ที่ระดับความเชื่อมั่น 90%**

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

จะได้ว่า วิธีการประมาณให้ค่าระดับความเชื่อมั่นไม่ต่ำกว่าระดับความเชื่อมั่นที่กำหนดถ้า $\hat{c}$ มีค่าอยู่ในช่วง

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

หรือมีค่าอยู่ในช่วง

$$(0.8890\ ,1)$$

**2. ที่ระดับความเชื่อมั่น 95%**

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

จะได้ว่า วิธีการประมาณให้ค่าระดับความเชื่อมั่นไม่ต่ำกว่าระดับความเชื่อมั่นที่กำหนดถ้า $\hat{c}$ มีค่าอยู่ในช่วง

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

หรือมีค่าอยู่ในช่วง

$$(0.9420\ ,1)$$

**3. ที่ระดับความเชื่อมั่น 99%**

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

จะได้ว่า วิธีการประมาณให้ค่าระดับความเชื่อมั่นไม่ต่ำกว่าระดับความเชื่อมั่นที่กำหนดถ้า $\hat{c}$ มีค่าอยู่ในช่วง

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

หรือมีค่าอยู่ในช่วง

$$(0.9864\ ,1)$$

นั่นคือค่าระดับความเชื่อมั่นที่ได้จากการทดลองจะต้องมีค่าไม่ต่ำกว่า 0.8890, 0.9420 และ 0.9864 ที่ระดับความเชื่อมั่น 90%, 95% และ 99% ตามลำดับ จึงจะถือว่าวิธีการประมาณนั้นให้ค่าระดับความเชื่อมั่นไม่ต่ำกว่าระดับความเชื่อมั่นที่กำหนด และจะนำเฉพาะวิธีการประมาณที่ให้ค่าระดับความเชื่อมั่นไม่ต่ำกว่าเกณฑ์ที่กำหนดนี้ไปทำการเปรียบเทียบค่าความยาวเฉลี่ยของค่าประมาณแบบช่วงต่อไป โดยจะทำการพิจารณาเปรียบเทียบค่าความยาวเฉลี่ยของของค่าประมาณแบบช่วงที่ได้ ว่าวิธีการประมาณใดสามารถให้ค่าความยาวเฉลี่ยของค่าประมาณแบบช่วงต่ำที่สุด จะถือว่าวิธีการประมาณนั้นให้ช่วงประมาณที่เหมาะสมที่สุดสำหรับสถานการณ์นั้น ๆ



## 2.10 ผลงานวิจัยที่เกี่ยวข้อง

สำหรับผลงานวิจัยที่เกี่ยวข้องกับการประมาณค่าสัดส่วนประชากรแบบช่วง เริ่มจาก Clopper และ Pearson (1934) ได้นำเสนอวิธีประมาณค่าสัดส่วนประชากรแบบช่วงซึ่งอยู่ในรูปของการแจกแจงเอฟ กรณีที่ตัวอย่างมีขนาดเล็ก พบว่าช่วงที่ได้สามารถให้ค่าระดับความเชื่อมั่นจากการทดลองไม่ต่ำกว่าเกณฑ์ที่กำหนด ต่อมามีผู้วิจัยหลายท่านได้ทำการศึกษา {Anderson และ Burstien (1967); Casella (1986) } พบว่าวิธีประมาณดังกล่าวจะประสบปัญหาในการคำนวณค่าเมื่อขนาดตัวอย่างเพิ่มขึ้น จึงทำให้มีนักวิจัยหลายท่านพยายามศึกษาหาวิธีที่จะใช้ในการประมาณค่า ดังเช่น

Ghosh (1979) ได้ทำการศึกษาเปรียบเทียบการประมาณค่าสัดส่วนประชากรแบบช่วงระหว่างวิธีปกติ วิธีแปลงแบบอาร์คไซน์ และวิธีสคอร์ ที่ระดับความเชื่อมั่น 95% และ 99% ขนาดตัวอย่างมีค่าเท่ากับ 15, 20, 30, 50, 100 และ 200 และค่าสัดส่วนประชากร 0.01, 0.05, 0.10(0.10)0.90, 0.95 และ 0.99 ผลจากการวิจัยสรุปได้ว่าวิธีสคอร์ สามารถให้ค่าระดับความเชื่อมั่นจากการทดลองไม่ต่ำกว่าเกณฑ์ที่กำหนด เมื่อค่าสัดส่วนประชากรมีค่าอยู่ในช่วง $0.01 \le p \le 0.99$ ถึงแม้ว่าขนาดตัวอย่างจะมีค่าเท่ากับ 20 วิธีปกติจะให้ค่าระดับความเชื่อมั่นจากการทดลองไม่ต่ำกว่าเกณฑ์ที่กำหนด เมื่อ $p > 0.20$ ถึงแม้ว่าจะใช้ขนาดตัวอย่างมากถึง 200 แล้วก็ตาม สำหรับวิธีแปลงแบบอาร์คไซน์ยังไม่มีข้อสรุปที่ชัดเจนแต่มีแนวโน้มไปทางวิธีสคอร์ มากกว่าวิธีปกติ

Chen (1990) ได้นำเสนอวิธีการประมาณค่าสัดส่วนประชากรแบบช่วง โดยนำหลักการของเบย์ ซึ่งถ้ากำหนดให้ $\beta = (z_{1-\alpha/2})^2/2$ พบว่าคือวิธีสคอร์ มาทำการศึกษาเปรียบเทียบการประมาณค่ากับวิธีปกติ และวิธีแปลงแบบอาร์คไซน์ ที่ระดับความเชื่อมั่น 95% ขนาดตัวอย่างมีค่าเท่ากับ 50, 100 และ 1,000 ค่าสัดส่วนประชากร 0.10, 0.05, 0.02, 0.01, 0.005, 0.001, 0.0005 และ 0.0001 ผลจากการวิจัยสรุปได้ว่าวิธีสคอร์ สามารถให้ค่าระดับความเชื่อมั่นไม่ต่ำกว่าเกณฑ์ที่กำหนด ที่สูงกว่าหรือเท่ากับวิธีปกติในทุกระดับขนาดตัวอย่าง และทุกระดับค่าสัดส่วนประชากรที่ทำการศึกษา