บทที่ 1

บทนำ

1.1 ที่มาและความสำคัญของปัญหา

การแจกแจงแบร์นูลลี (Bernoulli distribution) เป็นการแจกแจงที่อธิบายถึงสถานการณ์ของการทดลองใด ๆ ที่มีผลของการทดลองเป็น 2 ลักษณะคือ "ลักษณะที่สนใจ" หรือเกิดผลสำเร็จ (Success) และ "ลักษณะที่ไม่สนใจ" หรือไม่เกิดผลสำเร็จ (Failure) โดยมีค่าความน่าจะเป็นของการเกิด "ลักษณะที่สนใจ" เท่ากับ p และความน่าจะเป็นของการเกิด "ลักษณะที่ไม่สนใจ" เท่ากับ $1-p$ และเมื่อทำการทดลองแบร์นูลลี (Bernoulli trial) จำนวน n ครั้ง โดยที่การทดลองในแต่ละครั้งเป็นอิสระต่อกันและมีค่าความน่าจะเป็นของการเกิด "ลักษณะที่สนใจ" เท่ากับ p คงที่ตลอดการทดลองจะได้ว่าจำนวนครั้งของ "ลักษณะที่สนใจ" ที่เกิดขึ้นจากการทดลองทั้ง n ครั้ง มีการแจกแจงแบบทวินาม (Binomial distribution) ที่มีพารามิเตอร์ n และ p นั่นคือ

ถ้ากำหนดให้ $X_1, X_2, \ldots, X_n$ คือตัวแปรสุ่มที่มีการแจกแจงแบบแบร์นูลลีที่เป็นอิสระต่อกัน โดยมีพารามิเตอร์คือ p จะได้ว่า $Y = \sum_{i=1}^{n} X_i$ คือตัวแปรสุ่มที่มีการแจกแจงแบบทวินามโดยมีพารามิเตอร์คือ n และ p

สำหรับการอนุมานที่เกี่ยวกับพารามิเตอร์ p หรือที่เรียกว่า ค่าสัดส่วนประชากร (Population Proportion) ที่เป็นการประมาณค่า อาจเป็นการประมาณแบบจุด (Point estimation) หรือการประมาณแบบช่วง (Interval estimation) ในการประมาณค่าแบบจุดจะได้ว่า ตัวประมาณภาวะน่าจะเป็นสูงสุด (maximum likelihood estimator) คือ $\hat{p} = Y/n$ มีคุณสมบัติไม่เอนเอียงและมีความแปรปรวนต่ำสุดอย่างเอกรูป (Uniformly Minimum Variance Unbiased Estimator : UMVUE) สำหรับการประมาณค่าแบบช่วงสามารถประมาณได้โดยวิธีปกติ (Normal method) ซึ่งจากทฤษฎีบทลิมิตเข้าสู่ส่วนกลาง (The Central Limit Theorem) ได้ว่าถ้าจำนวนครั้งของการทดลอง n มีขนาดใหญ่ ค่าสัดส่วนตัวอย่าง (Sample Proportion: $\hat{p}$) จะมีแนวโน้มลู่เข้าสู่การแจกแจงแบบปกติที่มีค่าเฉลี่ย p และความแปรปรวน $p(1-p)/n$ ซึ่งจะได้ค่าประมาณแบบช่วงอยู่ในรูปแบบ

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

ที่ระดับความเชื่อมั่น $100(1-\alpha)\%$ โดยประมาณ

สำหรับการวิจัยครั้งนี้ผู้วิจัยสนใจวิธีการประมาณค่าสัดส่วนประชากรแบบช่วง ซึ่งพบว่าสามารถประมาณได้หลายวิธีและแต่ละวิธีมีข้อแนะนำสำหรับการใช้ที่แตกต่างกัน ดังเช่น

เมื่อ $n \to \infty$ **และ** $p \to \frac{1}{2}$ เป็นที่ทราบโดยทั่วกันว่าสามารถใช้วิธีปกติประมาณการแจกแจงทวินามได้ดี แต่บางครั้งอาจไม่เป็นไปตามข้อตกลงดังกล่าวนี้ และพบว่ามีข้อแนะนำจากนักวิชาการหลายท่านแตกต่างกัน เช่นแนะนำว่าควรใช้วิธีปกติประมาณการแจกแจงทวินามเมื่อ

- np และ $n(1-p)$ มากกว่า 5 { Aczel (1993); Anderson, Sweeney and Williams (1994); Creighton (1994) }
- $n \rangle 9\max\{(1-p)/p, p/(1-p)\}$ { Larson (1995) }
- $np(1-p) \rangle 3$ { Olkin (1994) }
- $np(1-p) \rangle 9$ { Ross (1994) }

เห็นได้ว่ายังไม่มีกฎเกณฑ์เป็นหลักที่ชัดเจนว่า ควรจะใช้วิธีปกติประมาณการแจกแจงทวินามเมื่อใด จึงอาจก่อให้เกิดความสับสนในการปฏิบัติงานจริง

เมื่อ $n \to \infty$ **และ** $p \to 0$ สามารถประมาณการแจกแจงแบบทวินามโดยใช้วิธีปัวส์ซอง (Poisson method) ซึ่งจะพบค่าประมาณแบบช่วงอยู่ในรูป

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

อย่างไรก็ตาม ก็ยังมีปัญหาเช่นเดียวกันกับการประมาณโดยใช้วิธีปกติ เนื่องจากยังมีข้อคิดเห็นที่ไม่ตรงกันในการพิจารณาว่าเมื่อใดจึงควรใช้วิธีปัวส์ซองประมาณการแจกแจงทวินาม เช่นใช้เมื่อ

- $n \geq 100$ และ $p \leq 0.05$ { Mason (1982) }
- $n \geq 100$ และ $np \leq 10$ { Creighton (1994); Hogg and Tanis (1993); Freund (1992) }
- n มีค่ามาก p มีค่าน้อย และ $np \leq 20$ { Olkin (1994) }

ซึ่งอาจก่อให้เกิดความสับสนในการปฏิบัติงานจริง ว่าควรใช้วิธีปัวส์ซองประมาณการแจกแจงทวินามเมื่อใด

เมื่อ n มีขนาดเล็ก Johnson and Kotz (1970) แนะนำให้ประมาณค่าสัดส่วนประชากรโดยใช้วิธีการของ Clopper-Pearson ซึ่งจะพบในรูปของการแจกแจงแบบเอฟ (F distribution) ดังนี้

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

นอกจากวิธีการดังกล่าวข้างต้นพบว่ายังมีวิธีการประมาณค่าวิธีอื่น ๆ ที่น่าสนใจ ดังเช่นวิธีสคอร์ (Score method) ตัวประมาณนี้จะได้มาจากส่วนกลับของตัวสถิติที่ใช้สำหรับการทดสอบสมมติฐาน (Inverting a Test Statistics) โดยมีรูปแบบของค่าประมาณคือ

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

หรือวิธีแปลงแบบอาร์คไซน์ (Arcsine transformation method) ซึ่งถือเป็นค่าประมาณแบบช่วงอีกวิธีหนึ่งที่ใช้หลักการของทฤษฎีบทลิมิตเข้าสู่ส่วนกลาง เช่นเดียวกับวิธีปกติ มีรูปแบบของค่าประมาณคือ

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

จะเห็นได้ว่าเรามีวิธีประมาณค่าแบบช่วงสำหรับค่าสัดส่วนประชากรได้หลายวิธี ดังนั้นเพื่อหาวิธีที่เหมาะสำหรับการประมาณค่า จึงได้ทำการวิจัยเพื่อศึกษาเปรียบเทียบวิธีการประมาณค่าแบบช่วงเหล่านี้ เพื่อหาข้อสรุปในการเลือกใช้วิธีการประมาณที่เหมาะสมสำหรับแต่ละสถานการณ์ n และ p

1.2 วัตถุประสงค์ของการวิจัย

การวิจัยมีวัตถุประสงค์ดังนี้

1.2.1 ศึกษาวิธีการประมาณค่าแบบช่วงสำหรับค่าสัดส่วนประชากร ดังนี้

1.2.1.1 วิธีปกติ

1.2.1.2 วิธีแปลงแบบอาร์คไซน์

1.2.1.3 วิธีสคอร์

1.2.1.4 วิธีปัวส์ซอง

1.2.1.5 วิธีเอฟ

1.2.2 เพื่อหาข้อสรุปในการเลือกใช้วิธีการประมาณที่เหมาะสม ในแต่ละสถานการณ์ของค่า n และ p

1.3 สมมติฐานของการวิจัย

การประมาณค่าสัดส่วนประชากรแบบช่วงโดยใช้วิธีสคอร์ จะให้ค่าระดับความเชื่อมั่นไม่ต่ำกว่าระดับความเชื่อมั่นที่กำหนดและให้ค่าความยาวเฉลี่ยของตัวประมาณต่ำกว่าวิธีการประมาณแบบอื่น ๆ ที่ศึกษาเปรียบเทียบสำหรับทุกค่าของ n และ p

1.4 ข้อตกลงเบื้องต้น

ขีดจำกัดความเชื่อมั่นของค่าประมาณแบบช่วงทั้ง 5 วิธีมีดังนี้

1.4.1 วิธีปกติ

ขีดจำกัดความเชื่อมั่นล่าง (p_L) : $\hat{p} - Z_{1-\alpha/2}\sqrt{\dfrac{\hat{p}\ (1-\hat{p})}{n}}$

ขีดจำกัดความเชื่อมั่นบน (p_U) : $\hat{p} + Z_{1-\alpha/2}\sqrt{\dfrac{\hat{p}\ (1-\hat{p})}{n}}$

1.4.2 วิธีแปลงแบบอาร์คไซน์

ขีดจำกัดความเชื่อมั่นล่าง (p_L) : $\sin^2\left[\arcsin\sqrt{\hat{p}} - \dfrac{Z_{1-\alpha/2}}{2\sqrt{n}}\right]$

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

1.4.3 วิธีสคอร์

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

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

1.4.4 วิธีปัวส์ซอง

ขีดจำกัดความเชื่อมั่นล่าง (p_L) : $\frac{1}{2n}\chi^2_{2y,\alpha/2}$

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

1.4.5 วิธีเอฟ

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

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

สัญลักษณ์ที่ใช้กำหนดข้างต้นมีความหมายดังนี้

p คือค่าสัดส่วนประชากร

$\hat{p}$ คือค่าสัดส่วนตัวอย่าง มีค่าเท่ากับ Y/n

Y คือจำนวนครั้งของผลสำเร็จ (ในตัวอย่าง)

n คือขนาดตัวอย่าง

$Z_{1-\alpha/2}$ คือค่าของตัวแปรสุ่มปกติมาตรฐาน ที่ให้ค่าฟังก์ชันความน่าจะเป็นสะสมเท่ากับ $1-\alpha/2$

$\chi^2_{2y,\alpha/2}$ คือค่าของตัวแปรสุ่มไค-สแคว์ ที่ให้ค่าฟังก์ชันความน่าจะเป็นสะสมเท่ากับ $\alpha/2$ และมีองศาความเป็นอิสระเท่ากับ $2y$

$\chi^2_{2(y+1),1-\alpha/2}$ คือค่าของตัวแปรสุ่มไค-สแคว์ ที่ให้ค่าฟังก์ชันความน่าจะเป็นสะสมเท่ากับ $1-\alpha/2$ และมีองศาความเป็นอิสระเท่ากับ $2(y+1)$

$F_{2(n-y+1),2y,1-\alpha/2}$ คือค่าของตัวแปรสุ่มเอฟ ที่ให้ค่าฟังก์ชันความน่าจะเป็นสะสมเท่ากับ $1-\alpha/2$ และมีองศาความเป็นอิสระ $2(n-y+1)$ และ $2y$

$F_{2(y+1),2(n-y),1-\alpha/2}$ คือค่าของตัวแปรสุ่มเอฟ ที่ให้ค่าฟังก์ชันความน่าจะเป็นสะสมเท่ากับ $1-\alpha/2$ และมีองศาความเป็นอิสระ $2(y+1)$ และ $2(n-y)$

1.5 ขอบเขตการวิจัย

1.5.1 กำหนดขนาดตัวอย่าง (n) มีค่าตั้งแต่ 2 ถึง 50

1.5.2 กำหนดค่าสัดส่วนประชากร (p) มีค่า 0.01(0.01)0.09 และ 0.10(0.05)0.50

1.5.3 กำหนดระดับความเชื่อมั่น 3 ระดับ คือ 90%, 95% และ 99%

1.5.4 ข้อมูลที่ใช้ในการวิจัยนี้ได้มาจากการจำลองข้อมูลโดยใช้เทคนิคมอนติคาร์โล (Monte Carlo simulation technique) และเขียนโปรแกรมโดยใช้ภาษา FORTRAN 77 ทำการทดลองซ้ำ 2,000 ครั้ง ในแต่ละสถานการณ์ (ข้อ 1.5.1-1.5.3)

1.6 คำจำกัดความ

ระดับความเชื่อมั่น (Level of Confidence) หมายถึง ความน่าจะเป็นที่พารามิเตอร์ของประชากรจะมีค่าอยู่ในช่วงของค่าที่ประมาณได้

1.7 ประโยชน์ที่คาดว่าจะได้รับจากการวิจัย

1.7.1 ผลที่ได้จากการวิจัยจะเป็นแนวทางในการตัดสินใจเลือกใช้วิธีประมาณค่าสัดส่วนประชากรแบบช่วงให้เหมาะสมกับแต่ละสถานการณ์ (n , p)

1.7.2 เป็นแนวทางในการศึกษาวิจัยเพื่อเปรียบเทียบวิธีการประมาณแบบช่วง สำหรับวิธีประมาณแบบอื่นหรือการแจกแจงอื่น ๆ ต่อไป