บทที่ 1

บทนำ

1.1 ความสำคัญและความเป็นมาของปัญหา

ในการทดสอบสมมติฐานเกี่ยวกับค่าสัดส่วนของประชากรนั้น ทฤษฎีลิมิตสู่ส่วนกลาง (Central Limit Theorem) กล่าวว่า การแจกแจงของค่าสัดส่วนตัวอย่างสามารถประมาณได้ด้วยการแจกแจงแบบปกติเมื่อตัวอย่างมีขนาดใหญ่พอ ถ้าเราให้ตัวแปรสุ่มแบบทวินามหมายถึง X โดย $X \sim b(n,p)$ ดังนั้น X จะมีการแจกแจงแบบปกติ $N(np, np(1-p))$ โดยประมาณเมื่อ n มีขนาดใหญ่ หรือ $\hat{P} = X/n \sim N(p, p(1-p)/n)$ โดยประมาณเมื่อ n มีขนาดใหญ่ เราเรียกตัวสถิติ $\hat{P} = X/n$ ว่า สัดส่วนตัวอย่าง (Sample proportion) โดยมี p เป็นสัดส่วนประชากร (Population proportion)

ช่วงความเชื่อมั่น $100(1-\alpha)$ เปอร์เซนต์ของค่าสัดส่วนประชากร บนพื้นฐานของการประมาณด้วยการแจกแจงแบบปกติ เขียนได้เป็น 2 แบบ คือ

รูปแบบที่ 1 $\hat{p} \pm (Z_{1-\alpha/2} SE(\hat{p}))$ ซึ่งเป็นรูปแบบของการประมาณด้วยการแจกแจงแบบปกติอย่างง่ายที่เห็นโดยทั่วไป คือไม่ใช้ค่าปรับแก้ไขเพื่อความต่อเนื่อง CC (Continuity correction)

รูปแบบที่ 2 $\hat{p} \pm (Z_{1-\alpha/2} SE(\hat{p}) + CC)$ ซึ่งเป็นรูปแบบที่ใช้ค่าปรับแก้ไขเพื่อความต่อเนื่องด้วย

WALTER W. HAUCK และ SHARON ANDERSON (1986) ได้ทำการศึกษาเปรียบเทียบ วิธีการประมาณแบบช่วงสำหรับผลต่างระหว่างค่าสัดส่วนกรณีสองประชากร โดยใช้การประมาณด้วยการแจกแจงแบบปกติ ซึ่งช่วงความเชื่อมั่นมีรูปแบบเป็น

$\hat{\Delta} \pm (Z_{1-\alpha/2} SE(\hat{\Delta}) + CC)$ เมื่อ $\hat{\Delta} = \hat{p}_1 - \hat{p}_2$ เป็นตัวประมาณแบบจุดของ $\Delta = p_1 - p_2$ โดย $\hat{p}_i = x_i/n_i$ เขาทำการศึกษาเปรียบเทียบระหว่างวิธีที่ไม่ใช้ค่าปรับแก้ไขเพื่อความต่อเนื่อง กับวิธีที่ใช้ค่าปรับแก้ไขเพื่อความต่อเนื่องหลายวิธี โดยทำการศึกษาเปรียบเทียบ

ในลักษณะที่ผสมผสานกับค่าความคลาดเคลื่อนมาตรฐาน (SE($\hat{p}$)) 2 ค่า ผลการศึกษาของเขาให้ข้อสรุปว่า ช่วงความเชื่อมั่นที่ประมาณด้วยการแจกแจงแบบปกติเมื่อไม่ใช้ค่าปรับแก้ไขเพื่อความต่อเนื่อง จะให้ค่าระดับความเชื่อมั่นต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด อย่างไรก็ตาม HAUCK และ ANDERSON ไม่ได้กล่าวสรุปรวมไปถึงการประมาณแบบช่วงสำหรับค่าสัดส่วนประชากรเดียว

ดังนั้นผู้วิจัยจึงสนใจที่จะศึกษาเปรียบเทียบ วิธีประมาณค่าสัดส่วนประชากรเดียวแบบช่วง ระหว่างการใช้ค่าปรับแก้ไขเพื่อความต่อเนื่องและวิธีการหาช่วงอย่างง่ายซึ่งไม่ใช้ค่าปรับแก้ไขเพื่อความต่อเนื่องภายใต้การแจกแจงแบบปกติ ทั้งนี้ค่าปรับแก้ไขเพื่อความต่อเนื่องที่ผู้วิจัยสนใจนำมาศึกษา เป็นค่าที่นิยมใช้และปรากฏในหนังสือหลายเล่ม คือ ค่าปรับแก้ไขเพื่อความต่อเนื่องของเยตส์ (Yates's Correction for Continuity) ซึ่งมีค่าเท่ากับ 1/(2n) ทำการศึกษาในลักษณะที่ผสมผสานกับค่าความคลาดเคลื่อนมาตรฐาน 2 ค่า ดังเช่นการศึกษาของ HAUCK และ ANDERSON นอกจากนี้ ผู้วิจัยยังสนใจที่จะศึกษาเปรียบเทียบวิธีดังกล่าวข้างต้นกับอีกวิธีหนึ่ง ซึ่งเป็นวิธีที่ประมาณด้วยการแจกแจงแบบเอฟ ดังรายละเอียดในวัตถุประสงค์ของการวิจัยที่จะกล่าวต่อไป เพื่อจะได้มีข้อสรุปว่า การประมาณค่าสัดส่วนประชากรแบบช่วงด้วยวิธีการประมาณใดเป็นวิธีที่ดีที่สุด ในกลุ่มวิธีที่ผู้วิจัยนำมาศึกษาและเป็นแนวทางในการหาค่าประมาณแบบช่วงที่เหมาะสมต่อไป

1.2 วัตถุประสงค์ของการวิจัย

เพื่อเปรียบเทียบ วิธีการประมาณค่าสัดส่วนประชากรแบบช่วง ด้วยวิธี

1. วิธีการประมาณด้วยการแจกแจงแบบปกติ มีรูปแบบคือ

$$\text{ขีดความเชื่อมั่นล่าง (PL)} = \hat{p} - \{ Z_{1-\alpha/2}\, SE(\hat{p}) + CC \}$$

$$\text{ขีดความเชื่อมั่นบน (PU)} = \hat{p} + \{ Z_{1-\alpha/2}\, SE(\hat{p}) + CC \}$$

1.1 ใช้ $CC = 0$ และ $SE(\hat{p}) = \{ \hat{p}(1-\hat{p})/n \}^{1/2}$

1.2 ใช้ CC = 0 และ $SE(\hat{p}) = \{ \hat{p}(1-\hat{p})/(n-1) \}^{1/2}$

1.3 ใช้ $CC = (2n)^{-1}$ และ $SE(\hat{p}) = \{ \hat{p}(1-\hat{p})/n \}^{1/2}$

1.4 ใช้ $CC = (2n)^{-1}$ และ $SE(\hat{p}) = \{ \hat{p}(1-\hat{p})/(n-1) \}^{1/2}$

2. วิธีการประมาณด้วยการแจกแจงแบบเอฟ

$$PL = \frac{x}{x + (n - x + 1)F1}$$

$$PU = \frac{(x + 1)F2}{(x + 1)F2 + (n - x)}$$

เมื่อ $F1 = F[1-\alpha/2;\ 2(n - x + 1),\ 2x]$

$F2 = F[1-\alpha/2;\ 2(x + 1),\ 2(n - x)]$

การเปรียบเทียบจะเปรียบเทียบจากค่าระดับความเชื่อมั่นของช่วงความเชื่อมั่น และค่าความยาวเฉลี่ยของช่วงความเชื่อมั่นที่คำนวณจากแต่ละวิธีการประมาณ ที่ระดับค่าสัมประสิทธิ์ความเชื่อมั่น 3 ระดับ คือ 90%, 95% และ 99%

1.3 สมมติฐานของการวิจัย

1. ในกรณีตัวอย่างขนาดเล็ก ช่วงความเชื่อมั่นของค่าสัดส่วนประชากรที่ประมาณด้วยการแจกแจงแบบปกติและไม่ใช้ค่าปรับแก้ไขเพื่อความต่อเนื่อง จะให้ค่าระดับความเชื่อมั่นต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด

2. วิธีการประมาณด้วยการแจกแจงแบบปกติและใช้ค่าปรับแก้ไขเพื่อความต่อเนื่องจะให้ระดับความเชื่อมั่นใกล้เคียงกับวิธีที่ประมาณด้วยการแจกแจงแบบเอฟ

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

1.4 ข้อตกลงเบื้องต้น

กรณีการประมาณค่าสัดส่วนประชากรแบบช่วงด้วยการแจกแจงแบบปกติ มีรูปแบบเป็น $\hat{p} \pm (Z_{1-\alpha/2} SE(\hat{p}) + CC)$ โดยมี

$\hat{p}$ คือ ค่าสัดส่วนตัวอย่าง เท่ากับ x/n

x คือ ค่าของตัวแปรสุ่มแบบทวินาม(จำนวนครั้งของผลสำเร็จในตัวอย่าง)

n คือ ขนาดตัวอย่าง

$Z_{1-\alpha/2}$ คือ ค่าของตัวแปรสุ่มปกติมาตรฐาน ที่ให้ค่าของฟังก์ชันความน่าจะเป็นสะสมเท่ากับ $1-\alpha/2$

$SE(\hat{p})$ คือ ค่าความคลาดเคลื่อนมาตรฐานของ $\hat{p}$

CC คือ ค่าปรับแก้ไขเพื่อความต่อเนื่อง

กรณีการประมาณค่าสัดส่วนประชากรแบบช่วงด้วยการแจกแจงแบบเอฟ

ขีดความเชื่อมั่นล่าง (PL) $= \dfrac{x}{x + (n - x + 1)F1}$

ขีดความเชื่อมั่นบน (PU) $= \dfrac{(x + 1)F2}{(x + 1)F2 + (n - x)}$

ซึ่ง

x คือ จำนวนครั้งของผลสำเร็จในตัวอย่าง

$F1$ คือ ค่าตัวแปรสุ่มเอฟ ที่ให้ค่าของฟังก์ชันความน่าจะเป็นสะสมเท่ากับ $1-\alpha/2$ และมีค่าองศาอิสระ $[2(n - x + 1), 2x]$

$F2$ คือ ค่าตัวแปรสุ่มเอฟ ที่ให้ค่าของฟังก์ชันความน่าจะเป็นสะสมเท่ากับ $1-\alpha/2$ และมีองศาอิสระ $[2(x + 1), 2(n - x)]$

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

1.5 ขอบเขตของการวิจัย

1. กำหนดขนาดตัวอย่างที่ใช้คือ n มีค่าตั้งแต่ 2 ถึง 50

2. กำหนดค่า p มีค่าตั้งแต่ 0.05 ถึง 0.95 โดยค่าเพิ่มขึ้นทีละ 0.05 รวมเป็นการแปรค่า p ทั้งสิ้น 19 ค่า

3. กำหนดระดับค่าสัมประสิทธิ์ความเชื่อมั่น $(1-\alpha)100\%$ เท่ากับ 90%, 95% และ 99%

4. ในการวิจัยครั้งนี้สร้างแบบจำลองข้อมูล โดยใช้เทคนิคมอนติคาร์โลซีมูเลชัน (Monte Carlo Simulation Technique) เขียนโปรแกรมด้วยภาษา FORTRAN 77 ทำการทดลองซ้ำ 2,000 ครั้ง ในแต่ละสถานการณ์ของการทดลอง (ข้อ 1, 2 และ 3)

1.6 คำจำกัดความ

ก. สัมประสิทธิ์ความเชื่อมั่น (Confidence coefficient) หมายถึง ความน่าจะเป็นที่ช่วงสุ่มจะครอบคลุมค่าของพารามิเตอร์ในประชากร

ข. ช่วงความเชื่อมั่น (Confidence interval) หมายถึง ช่วงตัวอย่างที่คำนวณจากข้อมูลตัวอย่างหนึ่งชุดใด ๆ ซึ่งใช้ในการประมาณค่าพารามิเตอร์แบบช่วง

1.7 ประโยชน์ที่คาดว่าจะได้รับจากการวิจัย

1. ผลที่ได้จากการวิจัยนี้ จะให้แนวทางในการตัดสินใจเลือกใช้วิธีการประมาณค่าสัดส่วนประชากรแบบช่วง ในแต่ละระดับขนาดตัวอย่างและแต่ละระดับค่าสัดส่วนตัวอย่าง

2. เป็นแนวทางในการวิจัยเพื่อศึกษาเปรียบเทียบ วิธีการประมาณแบบช่วงสำหรับผลต่างระหว่างค่าสัดส่วนกรณีสองประชากร

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย