บทที่ 1

บทนำ

1.1 ที่มาและความสำคัญของปัญหา

ในการประมาณค่าพารามิเตอร์ของประชากรนั้น เราสามารถทำการประมาณค่าได้ 2 แบบ คือ การประมาณค่าแบบจุด (point estimation) และการประมาณค่าแบบช่วง (interval estimation) สำหรับการประมาณค่าแบบจุด เป็นการประมาณค่าพารามิเตอร์ด้วยค่าประมาณเพียงค่าเดียว เราไม่สามารถบอกได้ว่าค่าประมาณแบบจุดนั้นมีค่าเท่ากับค่าพารามิเตอร์หรือแตกต่างจากค่าพารามิเตอร์อย่างไร และไม่อาจกล่าวได้ว่าการประมาณค่าแบบจุดไม่มีความผิดพลาด เช่น ไม่อาจกล่าวว่า $\bar{X}$ มีค่าเท่ากับ μ เพียงแต่คาดว่า $\bar{X}$ จะมีค่าไม่ห่างไกลจาก μ มากนัก ในทำนองเดียวกันก็คาดว่า S^2 จะไม่ห่างไกลจาก σ^2 การที่เราไม่สามารถระบุระดับของความไม่แน่นอนเกี่ยวกับค่าประมาณนั้น ๆ ได้ จึงเป็นจุดอ่อนของการประมาณค่าแบบจุด วิธีการที่จะระบุระดับความไม่แน่นอนเกี่ยวกับค่าประมาณคือ การระบุความน่าจะเป็นของความผิดพลาด (probability of error) โดยใช้วิธีการประมาณค่าแบบช่วงซึ่งเป็นการประมาณค่าพารามิเตอร์ในลักษณะที่บอกช่วง ๆ หนึ่ง ที่จะครอบคลุมค่าจริงของพารามิเตอร์ เมื่อกำหนดความน่าจะเป็นของความผิดพลาด เราสามารถหาช่วงที่ครอบคลุมค่าจริงของพารามิเตอร์ด้วยความเชื่อมั่นระดับหนึ่ง

สำหรับในการวิจัยครั้งนี้ ผู้วิจัยสนใจศึกษาการประมาณค่าแบบช่วงของค่าพารามิเตอร์ที่เป็นค่าสัดส่วนประชากร บนพื้นฐานของการประมาณด้วยการแจกแจงแบบปกติ

การประมาณค่าแบบช่วงสำหรับสัดส่วนของประชากร โดยทฤษฎีลิมิตสู่ส่วนกลาง (Central Limit Theorem) ได้ว่า ถ้าตัวอย่างสุ่มมีขนาดใหญ่พอ การแจกแจงของสัดส่วนตัวอย่างสามารถประมาณได้ด้วยการแจกแจงปกติ ถ้าให้ X เป็นตัวแปรสุ่มแบบทวินาม นั่นคือ X = จำนวนผลสำเร็จทั้งหมดในการทดลองแบบแบร์นูลลีที่เป็นอิสระซึ่งกันและกัน n ครั้ง และเขียนได้เป็น $X \sim B(n,p)$ ดังนั้น X จะมีการแจกแจงแบบปกติ $N(np,np(1-p))$ โดยประมาณ ($X \overset{.}{\sim} N(np,np(1-p))$) เมื่อ n มีขนาดใหญ่ และถ้า $\hat{P} = X/n$ จะได้

$\hat{P} \sim N(p, p(1-p)/n)$ เราเรียกตัวสถิติ $\hat{P}$ ว่าสัดส่วนตัวอย่าง (Sample proportion) โดยมี p เป็นสัดส่วนประชากร (Population proportion) และตัวแปรสุ่ม $\dfrac{\hat{P} - p}{\{ p(1-p)/n \}^{1/2}} \sim N(0,1)$ เมื่อ n มีขนาดใหญ่ ดังนั้นรูปแบบหนึ่งซึ่งมักใช้กันทั่วไปของการประมาณค่าแบบช่วงสำหรับสัดส่วนประชากร p เมื่อกำหนดความน่าจะเป็นของความผิดพลาด α หรือ สัมประสิทธิ์ความเชื่อมั่น (confidence coefficient) $1-\alpha$ คือ

$$\hat{P} \pm z_{1-\alpha/2} \{ \hat{P}(1-\hat{P})/n \}^{1/2}$$

ซึ่ง $z_{1-\alpha/2}$ คือเปอร์เซ็นไทล์ที่ $(1-\alpha/2)100$ ของ $N(0,1)$ และเรียกช่วงที่คำนวณได้ว่าช่วงความเชื่อมัน $100(1-\alpha)\%$ สำหรับ p รูปแบบของวิธีการประมาณข้างต้นเป็นรูปแบบอย่างง่ายสามารถให้ค่าประมาณที่ดีหรือให้ค่าสัมประสิทธิ์ความเชื่อมั่นไม่ต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด เมื่อ n มีขนาดใหญ่ อย่างไรก็ตามในทางปฏิบัติการรวบรวมข้อมูลจากตัวอย่างบางครั้งไม่อาจหาข้อมูลได้จำนวนมาก เนื่องจากมีข้อจำกัดเกี่ยวกับเวลาและทรัพยากร หรือมีความจำเป็นที่ต้องใช้ข้อมูลเท่าที่มีอยู่ การประมาณค่าแบบช่วงสำหรับสัดส่วนประชากรโดยใช้วิธีการประมาณอย่างง่าย จึงอาจให้ช่วงความเชื่อมั่นที่มีค่าสัมประสิทธิ์ความเชื่อมั่นต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่ต้องการ ทำให้ช่วงที่หาได้จากวิธีการประมาณดังกล่าวไม่ได้คุณสมบัติตามที่คาดหวัง

ด้วยสาเหตุดังกล่าวจึงเป็นที่น่าสนใจ ในการหาวิธีการประมาณค่าแบบช่วงสำหรับสัดส่วนประชากรด้วยการแจกแจงแบบปกติที่เหมาะสมกว่าวิธีการประมาณอย่างง่าย จากการศึกษาของผู้วิจัยพบว่ามีวิธีที่น่าสนใจอีกหลายวิธี ดังนั้นในการวิจัยครั้งนี้ ผู้วิจัยสนใจที่จะศึกษาเปรียบเทียบวิธีการประมาณค่าแบบช่วงสำหรับสัดส่วนประชากรบนพื้นฐานของการประมาณด้วยการแจกแจงแบบปกติ 3 วิธีคือ

1. วิธีการประมาณอย่างง่าย (Classical Method)

ช่วงความเชื่อมั่น 100(1-α)% สำหรับ p : (PL,PU)

ขีดจำกัดความเชื่อมั่นล่าง (PL) คือ $\hat{P} - z_{1-\alpha/2}\{ \hat{P}(1-\hat{P})/n \}^{1/2}$

ขีดจำกัดความเชื่อมั่นบน (PU) คือ $\hat{P} + z_{1-\alpha/2}\{ \hat{P}(1-\hat{P})/n \}^{1/2}$

2. วิธีการประมาณด้วยรากของสมการกำลังสอง (Root of Quadratic Equation Method)

ช่วงความเชื่อมั่น 100(1-α)% สำหรับ p : (PL,PU)

ขีดจำกัดความเชื่อมั่นล่าง (PL) คือ

$$\frac{\hat{P} + (z_o)^2/2n - z_o \{ \hat{P}(1-\hat{P})/n + (z_o)^2/4n^2 \}^{1/2}}{1+(z_o)^2/n}$$

ขีดจำกัดความเชื่อมั่นบน (PU) คือ

$$\frac{\hat{P} + (z_o)^2/2n + z_o \{ \hat{P}(1-\hat{P})/n + (z_o)^2/4n^2 \}^{1/2}}{1+(z_o)^2/n}$$

เมื่อ $z_o = z_{1-\alpha/2}$

3. วิธีการประมาณด้วยตัวประมาณเบส์โดยเชน (Bayesian Estimation Method by Chen)

ช่วงความเชื่อมั่น 100(1-α)% สำหรับ p : (PL,PU)

ขีดจำกัดความเชื่อมั่นล่าง (PL) คือ

$$\frac{X+b}{n+2b} - (z_o/n^{1/2})\{ \left(\frac{X+b}{n+2b}\right)\left(1-\frac{X+b}{n+2b}\right) \}^{1/2}$$

ขีดจำกัดความเชื่อมั่นบน (PU) คือ

$$\frac{X+b}{n+2b} + (z_o/n^{1/2})\{ \left(\frac{X+b}{n+2b}\right)\left(1-\frac{X+b}{n+2b}\right) \}^{1/2}$$

เมื่อ $z_o = z_{1-\alpha/2}$, $b = 0.375$ [1]

1.2 วัตถุประสงค์ของการวิจัย

เพื่อเปรียบเทียบ วิธีการประมาณค่าแบบช่วงสำหรับสัดส่วนประชากรบนพื้นฐานของการประมาณด้วยการแจกแจงแบบปกติ ด้วยวิธี

1. วิธีการประมาณอย่างง่าย
2. วิธีการประมาณด้วยรากของสมการกำลังสอง
3. วิธีการประมาณด้วยตัวประมาณเบส์โดยเชน

การเปรียบเทียบจะทำการเปรียบเทียบด้วยค่าสัมประสิทธิ์ความเชื่อมั่นของช่วงความเชื่อมั่น และค่าความยาวเฉลี่ยของช่วงความเชื่อมั่นที่คำนวณได้จากแต่ละวิธี ที่ระดับความเชื่อมั่น 3 ระดับคือ 90%, 95% และ 99% และขนาดตัวอย่าง 1 ถึง 50

1.3 สมมติฐานของการวิจัย

ในกรณีค่าสัดส่วนประชากรมีค่าเข้าใกล้ 0 หรือ 1 วิธีการประมาณด้วยตัวประมาณเบส์โดยเชน จะให้ค่าสัมประสิทธิ์ความเชื่อมั่นไม่ต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด และให้ค่าความยาวเฉลี่ยของช่วงความเชื่อมั่นต่ำกว่าวิธีการประมาณอย่างง่าย และวิธีการประมาณด้วยรากของสมการกำลังสอง

---

[1] Johnson N.L. and Kotz S. Discrete Distributions. New York: John Wiley & Son, 1969.

1.4 ข้อตกลงเบื้องต้น

วิธีการประมาณค่าแบบช่วงสำหรับสัดส่วนประชากรทั้ง 3 วิธีคือ

1. วิธีการประมาณอย่างง่าย

รูปแบบของการประมาณค่าแบบช่วงสำหรับสัดส่วนประชากรคือ

$$\hat{P} \pm z_{1-\alpha/2}\{\hat{P}(1-\hat{P})/n\}^{1/2}$$

2. วิธีการประมาณด้วยรากของสมการกำลังสอง

รูปแบบของการประมาณค่าแบบช่วงสำหรับสัดส่วนประชากรคือ

$$\frac{\hat{P} + (z_o)^2/2n \pm z_o\{\hat{P}(1-\hat{P})/n + (z_o)^2/4n^2\}^{1/2}}{1+(z_o)^2/n}$$

3. วิธีการประมาณด้วยตัวประมาณเบส์โดยเชน

รูปแบบของการประมาณค่าแบบช่วงสำหรับสัดส่วนประชากรคือ

$$\frac{X+b}{n+2b} \pm (z_o/n^{1/2})\{(\frac{X+b}{n+2b})(1-\frac{X+b}{n+2b})\}^{1/2}$$

โดยมี

$\hat{P}$	คือ	สัดส่วนตัวอย่าง เท่ากับ X/n
X	คือ	ตัวแปรสุ่มแบบทวินาม (จำนวนครั้งของผลสำเร็จในตัวอย่าง)
n	คือ	ขนาดตัวอย่าง
$z_o = z_{1-\alpha/2}$	คือ	เปอร์เซ็นไทล์ที่ $(1-\alpha/2)100$ ของตัวแปรสุ่มปกติมาตรฐาน
b	คือ	ค่าคงที่ เท่ากับ 0.375

1.5 ขอบเขตของการวิจัย

1. กำหนดขนาดตัวอย่างที่ใช้คือ n มีค่าตั้งแต่ 1 ถึง 50
2. กำหนดค่า p มี 2 ระดับคือ
   ระดับ 1 p มีค่าตั้งแต่ 0.01 ถึง 0.09 โดยค่า p เพิ่มขึ้นทีละ 0.01
   ระดับ 2 p มีค่าตั้งแต่ 0.10 ถึง 0.50 โดยค่า p เพิ่มขึ้นทีละ 0.05
   รวมเป็นค่า p ทั้งสิ้น 18 ค่า
3. กำหนดระดับความเชื่อมั่น $100(1-\alpha)\%$ เท่ากับ 90%, 95% และ 99%
4. ในการวิจัยครั้งนี้ สร้างข้อมูลโดยใช้เทคนิคการจำลองแบบมอนติคาร์โล (Monte Carlo Simulation Technique) เขียนโปรแกรมด้วยภาษาฟอร์แทรน 77 (FORTRAN 77) โดยทำการทดลองซ้ำ 2,000 ครั้ง ในแต่ละสถานการณ์ของการทดลอง (ข้อ 1, 2 และ 3)

1.6 คำจำกัดความ

ก. สัมประสิทธิ์ความเชื่อมั่น (Confidence coefficient) หมายถึง ความน่าจะเป็นที่ช่วงสุ่มจะครอบคลุมค่าจริงของพารามิเตอร์

ข. ช่วงความเชื่อมั่น (Confidence interval) หมายถึง ช่วงค่าประมาณพารามิเตอร์ด้วยระดับสัมประสิทธิ์ความเชื่อมั่นที่กำหนด

1.7 ประโยชน์ที่คาดว่าจะได้รับจากการวิจัย

1. ผลที่ได้จากการวิจัยนี้ จะให้แนวทางในการเลือกใช้วิธีการประมาณค่าแบบช่วงสำหรับสัดส่วนประชากรบนพื้นฐานของการประมาณด้วยการแจกแจงแบบปกติ ที่แต่ละระดับขนาดตัวอย่างและระดับค่าสัดส่วนตัวอย่าง
2. เพื่อเป็นแนวทางในการศึกษาเปรียบเทียบกับวิธีการประมาณแบบอื่นต่อไป