บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

การอนุมานเชิงสถิติ (statistical inference) เป็นระเบียบวิธีทางสถิติที่สำคัญอย่างหนึ่งในการหาข้อสรุปเกี่ยวกับประชากร ในลักษณะของการประมาณค่าพารามิเตอร์ (parameter estimation) และการทดสอบสมมติฐาน (hypothesis testing) ซึ่งการทดสอบสมมติฐานจะเป็นการนำค่าสถิติที่ได้จากตัวอย่างมาคำนวณค่าของตัวสถิติที่สอดคล้องกับเรื่องที่สนใจศึกษาเพื่อใช้ในการตัดสินใจว่าควรจะปฏิเสธหรือยอมรับสมมติฐานนั้น ส่วนการประมาณค่าพารามิเตอร์จะเป็นการนำค่าสถิติที่ได้จากตัวอย่างไปประมาณค่าพารามิเตอร์ของประชากร

การประมาณค่าพารามิเตอร์จะสามารถแบ่งออกได้เป็นสองประเภท ได้แก่ การประมาณแบบจุด (point estimation) และการประมาณแบบช่วง (interval estimation) สำหรับการประมาณแบบจุด เป็นการประมาณค่าพารามิเตอร์ หรือ ฟังก์ชันพารามิเตอร์ ที่อาศัยค่าประมาณที่เป็นตัวเลขซึ่งคำนวณได้จากตัวอย่างเพียงค่าเดียวหรือจำนวนเดียว โดยคาดว่าตัวประมาณนี้ จะให้ค่าที่ใกล้เคียงกับค่าที่แท้จริงซึ่งไม่ทราบค่า ส่วนการประมาณแบบช่วง เป็นการประมาณพารามิเตอร์หรือฟังก์ชันพารามิเตอร์ ที่อาศัยค่าประมาณเป็นช่วงของตัวเลข ซึ่งมีความเชื่อมั่นว่าช่วงดังกล่าวนี้จะครอบคลุมค่าที่แท้จริงของพารามิเตอร์หรือฟังก์ชันพารามิเตอร์ โดยพิจารณาจากรูปการแจกแจงของฟังก์ชันที่สร้างขึ้น ซึ่งประกอบด้วยตัวประมาณค่าแบบจุดและพารามิเตอร์ แล้วอาศัยรูปการแจกแจงดังกล่าวพยายามหาช่วงปริภูมิของพารามิเตอร์ หรือ ฟังก์ชันพารามิเตอร์ ที่จะเป็นไปได้ โดยให้มีความเชื่อมั่นเท่ากับที่กำหนดไว้ ซึ่งสิ่งที่มีผลกระทบต่อความแคบหรือความกว้างของช่วง คือ สัมประสิทธิ์ความเชื่อมั่น การแจงของตัวสถิติ และขนาดตัวอย่าง ดังนั้นจึงควรเลือกใช้ตัวสถิติที่มีความเหมาะสมกับเรื่องที่สนใจศึกษาและสอดคล้องกับประชากรเพื่อให้ช่วงความเชื่อมั่นที่ได้มีความถูกต้องและน่าเชื่อถือมากยิ่งขึ้น การประมาณค่าแบบช่วงที่สำคัญแบบหนึ่งคือ การประมาณค่าแบบช่วงสำหรับค่าเฉลี่ยของประชากร โดยทั่วไปจะใช้ตัวสถิติ Z (Z -statistic) ในการสร้างช่วงความเชื่อมั่นสำหรับค่าเฉลี่ยของประชากรเมื่อทราบส่วนเบี่ยงเบนมาตรฐานของประชากร และตัวสถิติที (t-statistic) ในการสร้างช่วงความเชื่อมั่นสำหรับค่าเฉลี่ยของประชากรเมื่อไม่ทราบส่วนเบี่ยงเบนมาตรฐานของประชากร โดยจะประมาณค่าส่วนเบี่ยงเบนมาตรฐานของประชากรด้วยส่วนเบี่ยงเบนมาตรฐานของตัวอย่าง การใช้ตัวสถิติดังกล่าวจะอยู่ภายใต้ข้อตกลงเบื้องต้น คือ ประชากรมีการแจกแจงแบบปกติ แต่ในบางครั้งเราจะพบว่าข้อมูล

หรือประชากรที่เราสนใจศึกษาไม่ได้มีการแจกแจงแบบปกติ ซึ่งการแจกแจงเหล่านี้มีอยู่หลายแบบและพบได้ในหลายๆเรื่อง

การแจกแจงที่ไม่ใช่แบบปกติที่มักใช้กันบ่อย คือ การแจกแจงแบบล็อกนอร์มอล (Log-normal distribution) โดยจะพบประชากรที่มีการแจกแจงแบบล็อกนอร์มอลได้ในหลายๆเรื่อง เช่น ระดับความสูงของน้ำในแม่น้ำ ระดับความสั่นสะเทือนของแผ่นดินไหว รายได้ของแต่ละบุคคล ข้อมูลทางการแพทย์ ได้แก่ ระยะเวลาในการฟักตัวของโรค ข้อมูลด้านการประกันภัย ได้แก่ ค่าสินไหมทดแทน เป็นต้น

ในกรณีที่ขนาดตัวอย่างมีขนาดใหญ่ก็จะสามารถประมาณได้ว่าข้อมูลที่ได้มีการแจกแจงใกล้เคียงกับการแจกแจงแบบปกติโดยอาศัยทฤษฎีค่าจำกัดสู่ส่วนกลาง (The Central Limit Theorem) แต่ในกรณีที่ขนาดตัวอย่างมีขนาดเล็กซึ่งอาจเป็นเพราะการเก็บข้อมูลทำได้ยาก ดังนั้นการใช้ตัวสถิติทีดังกล่าวในการสร้างช่วงความเชื่อมั่นสำหรับค่าเฉลี่ยอาจจะทำให้ผลสรุปที่ได้มีความผิดพลาด

ให้ X เป็นตัวแปรสุ่มที่มีการแจกแจงแบบล็อกนอร์มอลที่มีพารามิเตอร์สองตัว (μ , σ^2) ซึ่งฟังก์ชั่นความหนาแน่นของ X จะอยู่ในรูปของ

$$f(x)=\frac{1}{x\sigma\sqrt{2\pi}}\exp\left[-\frac{\{\ln(x)-\mu\}^2}{2\sigma^2}\right] \qquad ,x>0,\quad \sigma^2>0$$

และ Y = ln X มีการแจกแจงแบบปกติ ที่มี ค่าเฉลี่ย μ และ ความแปรปรวน σ^2 จะเห็นว่า X ต้องมีค่าเป็นบวกเสมอ ให้ θ เป็นค่าเฉลี่ยของ X ซึ่งอยู่ในรูปของฟังก์ชันของ μ และ σ ดังนี้

$$\theta=\exp\left(\mu+\frac{\sigma^2}{2}\right)$$

กำหนด $\bar{Y}=\frac{1}{n}\sum_{i=1}^{n}Y_i$ และ $S^2=\frac{1}{n-1}\sum_{i=1}^{n}\left(Y_i-\bar{Y}\right)^2$

ซึ่งวิธีการหาค่าประมาณแบบช่วงของ θ อย่างง่าย (naive method) ทำได้โดย ขั้นแรกสร้างช่วงความเชื่อมั่นสำหรับ μ จะได้ $\bar{Y}\pm Z_{\alpha/2}\left(S/\sqrt{n}\right)$ จากนั้นก็ใส่ ฟังก์ชันเอกซ์โปเนนเชียล (exponential) ที่ค่าของช่วงที่ได้ จะได้ $\exp\left(\bar{Y}\pm Z_{\alpha/2}\left(S/\sqrt{n}\right)\right)$ เป็นช่วงความเชื่อมั่นสำหรับ θ โดยประมาณ แต่จะเห็นว่าวิธีนี้เป็นการหาช่วงความเชื่อมั่นสำหรับ e^{μ} ซึ่งเป็นค่ามัธยฐาน (median) ไม่ใช่สำหรับค่าเฉลี่ย $\theta=e^{\mu+\sigma^2}$ ซึ่งวิธีนี้เหมาะสมกับตัวอย่างขนาดเล็กและเมื่อ σ^2 มีค่าน้อยๆ แต่จะไม่เหมาะเมื่อ σ^2 มีค่ามากและเมื่อตัวอย่างมีขนาดใหญ่ขึ้น

ในปี ค.ศ. 1966 Patterson ได้เสนอวิธีที่ใช้การแปลงลิมิตความเชื่อมั่น $\mu_\alpha = \hat{\mu} + t_{n-1}(\alpha)(\hat{\sigma}^2/n)^{1/2}$ สำหรับหาค่าเฉลี่ยของ log X โดยที่ $t_{n-1}(\alpha)$ คือเปอร์เซนไทล์ที่ α ของการแจกแจงแบบที ที่มีองศาแห่งความเป็นอิสระเท่ากับ $n-1$ นั่นคือ $\zeta_\alpha = \exp\left(\mu_\alpha + \frac{1}{2}\hat{\sigma}^2\right)$ เพื่อแก้ปัญหาจากวิธีการประมาณอย่างง่าย แต่วิธีนี้ก็ให้ผลที่ไม่ดีเมื่อ σ^2 มีค่าเพิ่มขึ้น และ ตัวอย่างมีขนาดใหญ่

ในปี ค.ศ. 1974 Cox ได้นำตัวประมาณที่มีคุณสมบัติไม่เอนเอียงและมีความแปรปรวนต่ำสุดอย่างเอกรูป (Uniformly Minimum Variance Unbiased Estimator) สำหรับ ln θ และความแปรปรวนของตัวประมาณนี้มาสร้างช่วงความเชื่อมั่น โดยใช้ทฤษฎีลิมิตเข้าสู่ส่วนกลาง (Central Limit Theorem) ในการสร้างช่วงความเชื่อมั่นสำหรับ ln θ ซึ่งจะให้ค่าสัมประสิทธิ์ความเชื่อมั่นสูงเมื่อตัวอย่างมีขนาดใหญ่

ในปี ค.ศ.1988 Angus ได้เสนอวิธีแบบคอนเซอเวทีฟในการสร้างช่วงความเชื่อมั่น โดยใช้ Approximate Pivotal Statistics ในการสร้างช่วงความเชื่อมั่นสำหรับ ln θ ซึ่งวิธีนี้จะให้ค่าสัมประสิทธิ์ความเชื่อมั่นสูง

ในปี ค.ศ. 1994 Angus ได้นำวิธีการบูทสแทร็พ (bootstrap) มาสร้างช่วงความเชื่อมั่น โดยใช้วิธี พาราเมตริก ทีเปอร์เซนไทล์ บูทสแทร็พ (parametric t-percentile bootstrap method) มาประยุกต์กับ Approximate Pivotal Statistics ในการสร้างช่วงความเชื่อมั่นสำหรับ ln θ ซึ่งวิธีการนี้ก็จะให้ค่าสัมประสิทธิ์ความเชื่อมั่นสูงเช่นกัน

วีรวรรณ ศักดาจิวะเจริญ (2544) ได้ทำการศึกษาเปรียบเทียบวิธีการประมาณค่าแบบช่วงสำหรับค่าเฉลี่ยของประชากรที่มีการแจกแจงแบบเบ้ขวา ซึ่งรวมการแจกแจงแบบล็อกนอร์มอลด้วย โดยมีวิธีการประมาณแบบช่วง 4 วิธี คือ วิธีการประมาณค่าแบบช่วงด้วยตัวสถิติที ตัวสถิติของจอห์นสัน ตัวสถิติของฮอลล์ และ ตัวสถิติของเชน ณ ระดับความเบ้ต่างๆ พบว่า ค่าสัมประสิทธิ์ความเชื่อมั่นจากการทดลองกรณีที่ใช้บูทสแทร็พในการหาช่วงความเชื่อมั่นมีค่าสูงกว่ากรณีที่ไม่ใช้บูทสแทร็พ

จากที่กล่าวมาข้างต้น และด้วยเหตุที่ยังไม่พบงานวิจัยที่เกี่ยวกับการเปรียบเทียบประสิทธิภาพของการประมาณช่วงความเชื่อมั่นสำหรับค่าเฉลี่ยของการแจกแจงแบบล็อกนอร์มอลด้วยวิธีการทั้ง 3 วิธีพร้อมกัน คือ

1. วิธีการประมาณของคอกซ์ (Cox's method)
2. วิธีการประมาณแบบคอนเซอเวทีฟ (Conservative method)
3. วิธีการประมาณแบบพาราเมตริกบูทสแทร็พ (Parametric bootstrap method)

ผู้วิจัยจึงสนใจที่จะทำการศึกษาการประมาณช่วงความเชื่อมั่นสำหรับค่าเฉลี่ยของการแจกแจงแบบล็อกนอร์มอล (log-normal distribution) เพื่อเปรียบเทียบวิธีการประมาณช่วงความเชื่อมั่นจาก 3 วิธีดังกล่าว เพื่อศึกษาหาข้อสรุปในการนำวิธีการประมาณแต่ละวิธีไปใช้ในแต่ละสถานการณ์ต่อไป

1.2 วัตถุประสงค์

การวิจัยมีวัตถุประสงค์ดังนี้

1.2.1 ศึกษาวิธีการประมาณค่าแบบช่วงสำหรับค่าเฉลี่ยของการแจกแจงแบบล็อกนอร์มอล ด้วยวิธีการ 3 วิธีคือ

วิธีที่ 1 วิธีการประมาณของคอกซ์ (Cox's method)

วิธีที่ 2 วิธีการประมาณแบบคอนเซอเวทีฟ (Conservative method)

วิธีที่ 3 วิธีการประมาณแบบพาราเมตริกบูทสแทร็พ (Parametric bootstrap method)

1.2.2 ศึกษาเปรียบเทียบวิธีการประมาณทั้ง 3 วิธี ด้วยค่าสัมประสิทธิ์ความเชื่อมั่น และความยาวเฉลี่ยของช่วงที่คำนวณได้จากแต่ละวิธี ที่สัมประสิทธิ์ความเชื่อมั่น 3 ระดับ คือ 0.90, 0.95 และ 0.99

1.2.3 เพื่อหาข้อสรุปในการเลือกใช้วิธีการประมาณช่วงที่เหมาะสม ในแต่ละสถานการณ์

1.3 สมมติฐานของการวิจัย

เมื่อตัวอย่างมีขนาดเล็กการประมาณค่าแบบช่วงโดยวิธีการประมาณแบบพาราเมตริกบูทสแทร็พ จะให้ค่าสัมประสิทธิ์ความเชื่อมั่นไม่ต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด และให้ค่าความยาวเฉลี่ยของช่วงความเชื่อมั่นสั้นกว่าวิธีการประมาณแบบอื่น

เมื่อตัวอย่างมีขนาดใหญ่การประมาณค่าแบบช่วงโดยวิธีการประมาณของคอกซ์จะให้ค่าสัมประสิทธิ์ความเชื่อมั่นไม่ต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนดและให้ค่าความยาวเฉลี่ยของช่วงความเชื่อมั่นสั้นกว่าวิธีการประมาณแบบอื่น

1.4 ขอบเขตของการวิจัย

ในการวิจัยครั้งนี้จะศึกษาเปรียบเทียบวิธีการประมาณค่าแบบช่วงสำหรับค่าเฉลี่ยของการแจกแจงแบบล็อกนอร์มอล โดยที่

1. ให้ X เป็นตัวแปรสุ่มที่มีการแจกแจงแบบล็อกนอร์มอลที่มีพารามิเตอร์สองตัว (μ , σ^2) ซึ่งฟังก์ชั่นความหนาแน่นของ X จะอยู่ในรูปของ

$$f(x)=\frac{1}{x\sigma\sqrt{2\pi}}\exp\left[-\frac{\{\ln(x)-\mu\}^2}{2\sigma^2}\right] \qquad , x>0,\quad \sigma^2>0$$

ซึ่งมี ค่าเฉลี่ย (mean) = $\exp\left(\mu+\frac{\sigma^2}{2}\right)$ และ

ความแปรปรวน (Variance) = $\exp(2\mu+\sigma^2)\{\exp(\sigma^2)-1\}$

ค่าสัมประสิทธิ์การกระจาย (coefficient of variation) ซึ่งแทนด้วย C.V. โดย

$$\text{C.V.} = \frac{\sqrt{\exp(2\mu+\sigma^2)\{\exp(\sigma^2)-1\}}}{\exp\left(\mu+\frac{\sigma^2}{2}\right)}$$

$$= \sqrt{\exp(\sigma^2)-1}$$

โดย พารามิเตอร์ของการแจกแจงแบบล็อกนอร์มอลจะกำหนดจาก C.V. ซึ่งจากสูตร จะเห็นว่า μ เป็นค่าใดๆ ในการวิจัยครั้งนี้จึงกำหนดค่า μ เป็น 2 ค่าคือ 1 และ 3 ดังนั้นจะได้ค่าพารามิเตอร์ที่ใช้ในการวิจัย ดังนี้

C.V.(%)	μ	σ^2
10	1,3	0.0100
50	1,3	0.2232
100	1,3	0.6932
150	1,3	1.1787
200	1,3	1.6095
250	1,3	1.9811
300	1,3	2.3026

2. กำหนดขนาดตัวอย่างที่ใช้ในการวิจัย มีค่าตั้งแต่ 5 ถึง 50
3. กำหนดค่าสัมประสิทธิ์ความเชื่อมั่น เท่ากับ 0.90 , 0.95 และ 0.99
4. การวิจัยครั้งนี้จะสร้างแบบจำลองข้อมูล โดยใช้เทคนิคมอนติคาร์โล (Monte Carlo Technique) ทำการทดลองซ้ำ 1,000 ครั้ง และใช้ขนาดตัวอย่างสำหรับวิธีบูทสแทร็พ (Bootstrap method) เท่ากับ 2,000

1.5 เกณฑ์ที่ใช้ในการประเมินประสิทธิภาพของช่วงความเชื่อมั่น

หลักเกณฑ์ในการพิจารณาเปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าแบบช่วงสำหรับค่าเฉลี่ยของการแจกแจงแบบล็อกนอร์มอลมีดังนี้

1) เปรียบเทียบค่าสัมประสิทธ์ความเชื่อมั่น (Confidence Coefficient) ที่คำนวณได้จากแต่ละวิธีการประมาณกับค่าสัมประสิทธิ์ที่กำหนด ในการตรวจสอบว่าวิธีการใดให้ค่าสัมประสิทธิ์ความเชื่อมั่นจากการทดลองไม่ต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนดได้หรือไม่นั้น ผู้วิจัยอาศัยการทดสอบสมมติฐานโดยใช้ตัวสถิติ Z ที่ระดับนัยสำคัญ 0.05 (รายละเอียดอยู่ในหัวข้อที่ 2.4)

2) ค่าความยาวเฉลี่ยของช่วงความเชื่อมั่นที่ต่ำกว่าเป็นวิธีการที่มีประสิทธิภาพมากกว่า ทั้งนี้ในการเปรียบเทียบค่าความยาวเฉลี่ยของช่วงความเชื่อมั่น จะเปรียบเทียบเฉพาะในกรณีวิธีการประมาณที่ให้ค่าสัมประสิทธิ์ความเชื่อมั่นจากการทดลองไม่ต่ำกว่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด

1.6 ขั้นตอนในการดำเนินการวิจัย

1. สร้างข้อมูลที่ใช้ในการวิจัย โดยใช้เทคนิคมอลติคาร์โลตามขนาดตัวอย่างที่กำหนด
2. ประมาณค่าพารามิเตอร์แบบจุดและความแปรปรวนของตัวประมาณค่า
3. คำนวณช่วงการประมาณค่าที่ต้องการศึกษาทั้ง 3 วิธี
4. คำนวณค่าสัมประสิทธิ์ความเชื่อมั่นของช่วงที่ประมาณได้ในขั้นตอนที่ 3 ของแต่ละวิธี
5. คำนวณค่าความยาวเฉลี่ยของช่วงที่ประมาณได้
6. เปรียบเทียบค่าสัมประสิทธิ์ความเชื่อมั่นที่ได้ในขั้นตอนที่ 4 กับเกณฑ์ที่กำหนดด้วยการทดสอบสมมติฐาน
7. เปรียบเทียบค่าความยาวเฉลี่ยของช่วงที่ประมาณได้ในแต่ละวิธีที่ผ่านเกณฑ์ที่กำหนดในขั้นตอนที่ 6
8. สรุปผลการวิจัยในแต่ละสถานการณ์

1.7 ประโยชน์ของการวิจัย

ประโยชน์ที่คาดว่าจะได้รับในการวิจัยครั้งนี้ คือ

1. เพื่อเป็นแนวทางในการเลือกวิธีการประมาณช่วงความเชื่อมั่นสำหรับค่าเฉลี่ยของการแจกแจงแบบล็อกนอร์มอล ในสถานการณ์ต่างๆได้อย่างเหมาะสม
2. เพื่อเป็นแนวทางในการศึกษาและเลือกวิธีการประมาณช่วงความเชื่อมั่นสำหรับค่าเฉลี่ยของการแจกแจงแบบล็อกนอร์มอลโดยใช้วิธีการประมาณอื่นๆต่อไป
3. เพื่อเป็นแนวทางในการศึกษาและเลือกวิธีการประมาณช่วงความเชื่อมั่นสำหรับค่าเฉลี่ยของการแจงแจงแบบอื่นๆต่อไป